

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-022398

(43)Date of publication of application : 21.01.1997

(51)Int.Cl.

G06F 15/16

G06F 15/16

G06F 12/00

G06F 13/00

(21)Application number : 07-170796

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 06.07.1995

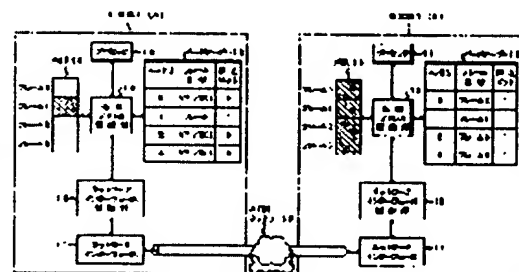
(72)Inventor : OKAMOTO TOSHIO
TSUDA YOSHIYUKI

(54) STORAGE SPACE MANAGEMENT METHOD, COMPUTER AND DATA TRANSFER METHOD IN DISTRIBUTED COMPUTER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To execute the data transfer of a shared space at a high speed by using a virtual channel at the time of transferring data between the shared area of one computer and the area of the other computer.

SOLUTION: This computer 1 is provided with a processor 15, a real memory 14, a virtual address conversion part 13, a page table 16, a network interface control part 12 and a network interface 11. Then, the call setting of ATM is performed between the areas in shared relation among the plural computers 1 and the virtual channel is laid. The identifier of the virtual channel is registered in the page table 16 of a virtual storage space management part and utilized at the time of actually transferring the data. In this case, when the access such as reference and write, to the shared area is actually generated and the transfer of the data between the shared computers 1 is required, since the virtual channel is laid, the data are transferred by using it.



(11)特許出願公開番号

特開平9-22398 ✓

(43)公開日 平成9年(1997)1月21日

(51)Int.Cl. ⁸		識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F	15/16	3 7 0		G 0 6 F	15/16 3 7 0 N
			7623-5B		12/00 5 4 5 M
	12/00	5 4 5	9460-5E		13/00 3 5 5
	13/00	3 5 5			15/16 4 0 0 M

審査請求 未請求 請求項の数5 OL (全 28 頁)

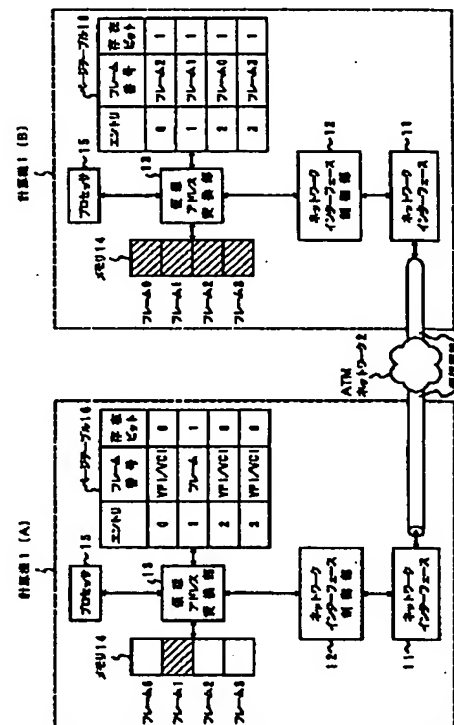
(21)出願番号	特願平7-170796	(71)出願人	000003078 株式会社東芝 神奈川県川崎市幸区堀川町72番地
(22)出願日	平成7年(1995)7月6日	(72)発明者	岡本 利夫 神奈川県川崎市幸区小向東芝町1番地 株 式会社東芝研究開発センター内
		(72)発明者	津田 悦幸 神奈川県川崎市幸区小向東芝町1番地 株 式会社東芝研究開発センター内
		(74)代理人	弁理士 鈴江 武彦

(54) 【発明の名称】 分散計算機システムにおける記憶空間管理方法、計算機及びデータ転送方法

(57) 【要約】

計算機側に複雑で効率の悪い通信プロトコル処理を必要とせず、計算機間での共有空間のデータ転送を高速に実行できる、効率のよい分散計算機システムにおける記憶空間管理方法を提供する。

【課題】 複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおける記憶空間管理方法であって、一方の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域と、他方の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化し、共有化した前記領域間ごとに前記仮想回線を専用に設定しておき、共有化した一方の計算機の前記領域と他方の計算機の前記領域との間でデータ転送を行なうにあたっては、当該共有化した領域間に専用に設定しておいた前記仮想回線を用いることを特徴とする。



1

【特許請求の範囲】

【請求項1】複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおける記憶空間管理方法であって、

一方の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域と、他方の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化し、
共有化した前記領域間ごとに前記仮想回線を専用に設定しておく、

共有化した一方の計算機の前記領域と他方の計算機の前記領域との間でデータ転送を行なうにあたっては、当該共有化した領域間に専用に設定しておいた前記仮想回線を用いることを特徴とする記憶空間管理方法。

【請求項2】複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおける計算機であって、

仮想記憶空間の仮想アドレスと対応する実記憶空間の物理アドレスまたはデータを所有する計算機にアクセスするための情報とを組にして登録したページテーブルを用いて仮想記憶空間を管理する手段と、

仮想記憶空間の全部または一部の領域と、他の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化するため、共有化する領域間ごとに前記仮想回線を専用に設定することを指示する手段と、前記仮想回線を専用に設定した前記領域に対応する前記ページテーブルのエントリに、前記データを所有する計算機にアクセスするための情報として、前記仮想回線の仮想回線識別子を登録する手段と、

共有化した前記領域間でデータ転送を行なうに先だって前記ページテーブルの該当するエントリに登録されている前記仮想回線識別子を参照し、得られた仮想回線識別子を指定し、当該共有化した領域間に専用に設定された前記仮想回線を用いてデータ転送を行なう手段とを備えたことを特徴とする計算機。

【請求項3】前記ページテーブルの前記仮想回線識別子を記入する項目には、デフォルト値として、データとこれを所有する計算機との対応を管理するネームサーバに接続するのに用いる仮想回線識別子を記入しておくことを特徴とする請求項2に記載の計算機。

【請求項4】複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおける計算機であって、

実記憶空間の全部または一部の領域と、他の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化するため、共有化する領域間ごとに前記仮想回線を専用に設定する手段と、

前記領域を共有化したときにまたは共有化した前記領域を処理実体がアクセスする前に、該領域間に専用に設定した前記仮想回線を用いてデータ転送を行なう手段とを

2

備えたことを特徴とする計算機。

【請求項5】複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおいて異なる計算機間で共有する仮想記憶空間の領域間でデータを転送するためのデータ転送方法であって、

一方の計算機が管理する仮想記憶空間の一部の領域と、他方の計算機が管理する仮想記憶空間の一部の領域とを共有化するとともに、共有する仮想記憶空間の1つの領域には複数の実メモリの領域を割当て、

10 共有化した前記仮想記憶空間の領域間ごとに前記仮想回線を専用に設定し、

前記仮想記憶空間の領域には、実際に対応させる実メモリの領域を順次切り替え、

前記仮想記憶空間の領域と現在対応していない実メモリの領域間で、前記仮想回線を用いてデータ転送を行い、データ転送の完了した実メモリの領域を、前記仮想記憶空間の領域に対応させることを特徴とするデータ転送方法。

【発明の詳細な説明】

20 【0001】

【発明の属する技術分野】本発明は、仮想記憶空間管理機構を有する複数の計算機システムを相互にATM（非同期転送モード）等のネットワークを介して接続してなる分散計算機システムにおける仮想記憶空間管理方法および計算機システムに関する。

【0002】

【従来の技術】近年、計算機性能の向上とネットワーク技術の進歩により、一連の処理を分割し、複数の計算機上でプログラムを動作させ、通信しながら協調して処理を進めていく分散処理が一般化してきている。

30 【0003】複数の計算機相互を接続して1つの分散計算機システムをつくる構成は、種々存在する。例えば、DEC社のVAX-Clusterは、通信チャネルを介して相互に直接結合した構成で実現されてきた。これは、専用の特別のハードウェアで実装されていた。また、高速通信を実現するため、計算機相互の距離は短くし、計算機相互を隣接させて設置することが必要であった。また、1つのシステムバス上に、複数のプロセッサを載せたハードウェア構成のマルチプロセッサシステムもある。

40 【0004】一方、近年のネットワーク技術の進歩により、高速なネットワーク装置が標準化され安価に提供されつつある。その代表例が、ATMを利用したネットワークである。このATMネットワークは、高速なスイッチ型交換機で実装され、WAN環境からLAN環境にいたるまで、幅広い範囲で利用できるように、通信プロトコルが定められている。例えば、呼設定、フロー制御、

50 【0005】このような高速ネットワークに既存の計算

3

機を相互に接続して構成した分散計算機システムがある。この方式によれば、上述のクラスタ接続やシステムバスによる方式と異なり、特別なハードウェアが必要ないので安価に実装できる利点がある。

【0006】しかし、従来の通信方式でATMを利用する場合、問題点が生じる。従来の通信方法では、計算機上の基本ソフトウェアであるOS（オペレーティングシステム）内等で、計算機用のネットワークプロトコル処理、例えばTCP/IPを用いて通信を行う。例えば、TCP/IP方式でこのATMを利用すると、ATMは単なる回線のごとく見なされ、ATMの特質を生かせない。計算機内ですべての通信プロトコル処理を行うので、処理が重く通信効率が悪くという欠点が生じる。

【0007】次に、従来の仮想記憶空間管理機構について説明する。従来の計算機においては、プログラム内に書かれた仮想アドレスは、ページテーブル等のアドレス変換機構によって、物理アドレスに変換され、変換された物理アドレスを用いて、計算機内の実メモリがアクセスされる。

【0008】以下、従来技術におけるアドレス変換機構について説明する。計算機内の実メモリは、使用効率を高めるため、512～8Kバイトのページと呼ばれる単位ごとに管理され使用される。個々のページは、「ページ・フレーム」と呼ばれる。

【0009】実メモリをアクセスするためには、ページを特定するために「フレーム番号」が必要であり、ページ内の位置を特定するために「オフセット」が必要である。従って、実メモリをアクセスするために用いる「物理アドレス」は、「フレーム番号」と「オフセット」の組からなる。

【0010】同様に、「仮想アドレス」は、「仮想フレーム番号」と「オフセット」の2つの部分に分割することができる。仮想アドレスから物理アドレスへの変換の際、仮想アドレスのオフセット部分は、変換されず、そのまま、物理アドレスのオフセット部分となる。

【0011】仮想フレーム番号については、ページテーブルの仮想フレーム番号に対応するページテーブル・エントリ内に、変換後の（ページ）フレーム番号が書かれており、ページテーブルを参照することによって、仮想フレーム番号から（ページ）フレーム番号に変換することができ、従って、仮想アドレスから物理アドレスへの変換を行うことができる。

【0012】なお、プログラムの実行に際し、仮想アドレスから物理アドレスへの変換は頻繁に行われるが、この変換の際に用いるページテーブルが実メモリ上に存在する場合、変換の都度、実メモリをアクセスしなければならない、変換に時間がかかる。これを解決するため、仮想アドレスから物理アドレスへの変換に用いたページテーブル・エントリを、プロセッサ内のTLB（Translation Lookaside Buffer）

4

と呼ばれるページテーブル・エントリ・キャッシュに格納する技術がある。これにより、その直後に同じ仮想フレームを変換する際には、実メモリをアクセスするよりも高速なTLBをアクセスすれば良いことになる。

【0013】次に、実メモリにロードされていない仮想フレームに属するデータをアクセスする場合について説明する。ページテーブル・エントリは、「存在ビット」を持つ。存在ビットが1の場合、対応するページ・フレームは実メモリ上にロードされていることを意味する。

【0014】一方、存在ビットが0の場合、対応するページ・フレームは実メモリ上にロードされていないことを意味する。存在ビットが0のページテーブル・エントリをアクセスすると、ページ・フォールトが発生し、OS（オペレーティング・システム）が、対応するページ・フレームを作成し、データを実メモリ上にロードしなければならない。

【0015】このような場合において、例えば、UNIX OSにおいて、ファイルシステムでのinode等のテーブルを参照することによって、該当するデータのハードディスク上での位置を特定し、必要な量のデータをハードディスクから実メモリへと転送する。

【0016】なお、上記データのハードディスク上での位置を効率よく知るために、上記データが格納されたハードディスク上でのセクタ番号等の情報を、ページテーブル・エントリに書き込むことによって、inode等のテーブルを参照するオーバーヘッドを無くすることができる。

【0017】次に、自分の計算機にないファイル等の資源をアクセスする場合について説明する。自分の計算機にないファイル等の資源のうち、以前に該資源を所有する計算機からネットワークを経由して転送した部分については、実メモリ上にキャッシュされたデータが残っている場合、対応するページテーブル・エントリの「存在ビット」は1であり、ページテーブル・エントリに書かれたフレーム番号を用いて、実メモリをアクセスし、該当データにアクセスすることができる。

【0018】しかし、今までに転送していない部分や、実メモリ上にキャッシュされたデータが残っていない場合は、対応するページテーブル・エントリの「存在ビット」は0であり、上記資源を所有する計算機からネットワークを経由して対応する部分を転送し実メモリ上にロードしなければならない。

【0019】この場合、今までに転送していない仮想フレームをアクセスするごとに、下記の参考文献～で解説されているような種々のテーブルを参照し、上記資源を所有する計算機を特定し、その計算機に対し、対応するデータの転送を要求するので、効率が悪いといった問題点があった。

【0020】仮想アドレスから物理アドレスへの変換については、例えば、「岩波講座マイクロエレクトロニ

10

20

30

40

50

クス 8 : V L S I コンピュータ I」 (元岡 達編) の p. 2 1 9 ~ p. 2 2 2 に詳しい。

T L B の動作については、例えば、「岩波講座マイクロエレクトロニクス 8 : V L S I コンピュータ I」 (元岡 達編) の p. 2 0 8 に詳しい。

ネットワークを経由して、他の計算機にアクセスする際に行われる手順については、例えば、「オペレーティングシステムの設計 I I X i n u によるインターネットワークの構築」 (啓学出版、ダグラス・カマー著) に詳しい。

【 0 0 2 1 】

【発明が解決しようとする課題】従来は、分散計算機システムにおいて計算機間が共有する記憶空間の間でデータ転送を行なうおうとすると、専用のハードウェアが必要となり、システム構成上、コストの点や柔軟性の点で不利であった。また、既存のネットワークを用いて複数の計算機を接続しただけでは、データ転送に時間がかかり、効率的な記憶空間の共有が困難であった。

【 0 0 2 2 】本発明は、上記事情を考慮してなされたものであり、計算機側に複雑で効率の悪い通信プロトコル処理を必要とせず、計算機間での共有空間のデータ転送を高速に実行できる、効率のよい分散計算機システムにおける記憶空間管理方法、計算機及びデータ転送方法を提供することを目的とする。

【 0 0 2 3 】

【課題を解決するための手段】本発明 (請求項 1) は、複数の計算機をネットワーク (例えば A T M ネットワーク) により提供される仮想回線で結合してなる分散計算機システムにおける記憶空間管理方法であって、一方の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域と、他方の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化し、共有化した前記領域間ごとに前記仮想回線を専用に設定しておき、共有化した一方の計算機の前記領域と他方の計算機の前記領域との間でデータ転送を行なうにあたっては、当該共有化した領域間に専用に設定しておいた前記仮想回線を用いることを特徴とする。

【 0 0 2 4 】本発明 (請求項 2) は、複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおける計算機であって、仮想記憶空間の仮想アドレスと対応する実記憶空間の物理アドレスまたはデータを所有する計算機 (オーナー計算機) にアクセスするための情報とを組にして登録したページテーブルを用いて仮想記憶空間を管理する手段と、仮想記憶空間の全部または一部の領域と、他の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化するため、共有化する領域間ごとに前記仮想回線を専用に設定することを指示する手段と、前記仮想回線を専用に設定した前記領域に対応する前記ページテーブルのエントリに、前記データを所有する計算機 (オ

ナー計算機) にアクセスするための情報として、前記仮想回線の仮想回線識別子を登録する手段と、共有化した前記領域間でデータ転送を行なうに先だって前記ページテーブルの該当するエントリに登録されている前記仮想回線識別子を参照し、得られた仮想回線識別子を指定し、当該共有化した領域間に専用に設定された前記仮想回線を用いてデータ転送を行なう手段とを備えたことを特徴とする、本発明 (請求項 3) は、上記発明 (請求項 2) において、前記ページテーブルの前記仮想回線識別子を記入する項目には、デフォルト値として、データとこれを所有する計算機との対応を管理するネームサーバに接続するのに用いる仮想回線識別子を記入しておくことを特徴とする。

【 0 0 2 5 】本発明 (請求項 4) は、複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおける計算機であって、実記憶空間の全部または一部の領域と、他の計算機が管理する仮想記憶空間または実記憶空間の全部または一部の領域とを共有化するため、共有化する領域間ごとに前記仮想回線を専用に設定する手段と、前記領域を共有化したときにまたは共有化した前記領域を処理実体がアクセスする前に、該領域間に専用に設定した前記仮想回線を用いてデータ転送を行なう手段とを備えたことを特徴とする。

【 0 0 2 6 】本発明 (請求項 5) は、複数の計算機をネットワークにより提供される仮想回線で結合してなる分散計算機システムにおいて異なる計算機間で共有する仮想記憶空間の領域間でデータを転送するためのデータ転送方法であって、一方の計算機が管理する仮想記憶空間の一部の領域と、他方の計算機が管理する仮想記憶空間の一部の領域とを共有化するとともに、共有する仮想記憶空間の 1 つの領域には複数の実メモリの領域を割当て、共有化した前記仮想記憶空間の領域間ごとに前記仮想回線を専用に設定し、前記仮想記憶空間の領域には、実際に対応させる (マッピングする) 実メモリの領域を順次切り替え、前記仮想記憶空間の領域と現在対応していない (マッピングされていない) 実メモリの領域間で、前記仮想回線を用いてデータ転送を行い、データ転送の完了した実メモリの領域を、前記仮想記憶空間の領域に対応させる (マッピングする) ことを特徴とする。

【 0 0 2 7 】 (作用) 本発明によれば、複数の計算機をネットワーク (例えば A T M ネットワーク) を介して結合した分散システムにおいて、仮想記憶空間または実記憶空間の全部または一部を共有化し、共有領域ごとに専用に設定しておいた仮想回線を利用して直接、高速なデータ転送を行なうことができるので、計算機側には複雑で効率の悪い通信プロトコル処理が必要とせず、効率的な記憶空間の共有が実現できる。つまり、その共有空間に対して参照や書き込み (read/write) が発生しても、その都度、呼設定不要で、直にデータ転送ができる。

【0028】また、呼設定、フロー制御、ふくそう制御、エラー検出、QoS管理などデータ転送中の制御はネットワーク側（例えばATM交換機側）で対応してくれるので、計算機側はその機能に不足する部分（例えばデータエラーにおける訂正や再送）だけ管理すればいいので、計算機側の処理が大幅に軽減する。

【0029】さらに、ATMネットワークのように高速なネットワークを利用することによって、ある程度の距離が離れた計算機どうしを結んで分散システムにできるので、構成の柔軟性が良い。

【0030】また、本発明によれば、他の計算機に属する資源にアクセスする場合、仮想アドレスを物理アドレスに変換する際に用いるページテーブルを参照するだけで、他の計算機へのコネクションを使用するために必要な仮想回線識別子（VPI/VCI値等）を得ることができる。

【0031】このため、他の計算機に属する資源のうち、必要な大きさのデータブロックの転送を要求する場合、従来のように種々のテーブルを参照することなく、ページテーブルから得られた仮想回線識別子を用いてデータ転送の要求セルを作成することができ、効率良く要求セルを作成することができる。

【0032】また、本発明によれば、前記ページテーブルの前記仮想回線識別子を記入する項目にデフォルト値として、データとこれを所有する計算機（オーナー計算機）との対応を管理するネームサーバに接続するのに用いる仮想回線識別子を記入しておけば、データを所有する計算機（オーナー計算機）を知る必要が生じた際に、該当するページテーブル・エントリに書き込まれている仮想回線識別子を用いて、速やかにネームサーバに問い合わせを行なうことができる。

【0033】また、本発明によれば、クライアント側の実記憶空間の全部または一部の領域と、サーバ計算機の仮想記憶空間または実記憶空間の全部または一部の領域とを共有化する場合、クライアント側の共有領域には、すでに実記憶空間（実メモリ）が割り当てられているので、ページフォルトは発生しない。この場合、クライアント側での実際のアクセスが発生した時にデータ転送するのではなく、仮想回線の設定時に、引き続いてデータ転送を行う（すなわちクライアント側のアクセスとは無関係である。）

したがって、ユーザプログラムやスレッド等の処理実体が仮想記憶空間にアクセスしたときには、既に実記憶空間（実メモリ）が割り当てられているので、ページフォルトが発生せず、ページフォルト処理が不要になり、即座にデータをアクセスできる。したがって、例えば、リアルタイム処理等の高速なユーザプログラムの実行が可能となる。

【0034】また、本発明によれば、クライアント計算機の仮想記憶空間の一部の領域と、サーバ計算機の仮想

記憶空間の一部の領域とを共有化するとともに、共有する仮想記憶空間の1つの領域には複数の実メモリの領域を割り当て、仮想記憶空間の領域と現在対応していない

（マッピングしていない）実記憶空間間で専用の仮想回線を用いてデータ転送を行うとともに、データ転送の完了した実メモリの領域を仮想記憶空間の領域に対応させる（マッピングする）ように、実メモリの領域を順次切り替えるので、計算機間で仮想記憶空間の領域の大きさ単位に連続して高速にデータを転送することができる。

10 【0035】

【発明の実施の形態】以下、図面を参照しながらこの発明の実施の形態について説明する。図1に、この実施の形態に係る分散システム全体の基本構成を示す。この実施の形態は、複数の計算機間の仮想記憶空間や実記憶空間（実メモリ）相互を共有化することで分散システムを構成し、その領域を相互に参照や書き込みを行うことで通信を行ない、その相互通信にATMを利用するものである。

【0036】図1中、2はATM方式によるスイッチ型高速LAN（Local Area Network）であるATMネットワークである。ATM交換機3に複数の

20 ホスト（計算機）1を結合して分散システムを構成している。

【0037】ATM交換機3には呼設定を制御する呼設定サーバ4が接続されている。呼設定サーバ4は、ATMネットワーク2中に1台だけ設けても良いし、複数台設けて機能分担させても良い。

【0038】各ホスト（計算機）は、1台または複数台のATM交換機3を経由して他のホスト（計算機）との間にコネクション（仮想回線）を設定することができる。ホスト1は、少なくとも1つのCPUとRAMなどのメモリ（実メモリ）を備える計算機であり、ネットワーク中で識別するためのアドレス等のような識別子を与える単位である。また、各ホスト1は、プログラム、データ、ファイルなどを記憶しておくディスク装置を備えているものとする。

【0039】なお、図1のように、ATMネットワーク2には、ファイル等の資源とこれを所有する計算機との対応等を管理するネームサーバ5を接続しても良い。ネームサーバ5は、ATMネットワーク2中に1台だけ設けても良いし、複数台設けて機能分担させても良い。また、ネームサーバ5は、計算機1の中に設けても良い。

【0040】次に、図2と図3を参照しながら、この実施の形態に係る計算機1の構成について説明する。図2は、この実施の形態に係る計算機1内部の要部構成を示す基本構成図であるとともに、2つの計算機1間をATMの仮想回線で接続した状態を示している。また、図3には、この計算機1内部の実際の物理構成の一例を示す。

50 【0041】図2に示すように、計算機1は、プロセッ

サ15、実メモリ14、仮想アドレス変換部13、ページテーブル16、ネットワーク・インターフェース制御部12、ネットワーク・インターフェース11を備えている。

【0042】図3中、17はシステムバスを、18はハードディスク制御部を、19はハードディスクを、151はバスインタフェース部を、152はプログラムを実行するエグゼキューション部を、153はTLB (Translation Lookaside Buffer) を示す。

【0043】図3に示すように、ページテーブル16は、物理構成上は実メモリ14上に格納され、ページテーブル16の一部がプロセッサ15内のTLB153の中にキャッシュされる。

【0044】仮想アドレス変換部13は、ページテーブル16を参照して、プログラム等の実行時にアクセスされた仮想アドレスを、これに対応する実メモリ14上のデータへの物理アドレスに変換する。なお、仮想アドレス変換部13は、説明上、図2では物理構成上プロセッサ15外に存在するよう示し、図3では省略しているが、実際には、プロセッサ15内に存在させても良いし、プロセッサ15外に存在させても良い。

【0045】次に、この実施の形態に係るオペレーティングシステム(OS)に関して説明を行う。計算機1には、分散システムを構成するために必要なオペレーティングシステム(OS)を搭載している。このOSは、プログラムやデータなどを配置するアドレス空間の管理方式として仮想記憶空間を提供する。

【0046】図4に、このOSの要部構成を示す。図4中、21はネットワーク共有メモリ管理部を、22は仮想記憶空間管理部を、23はネットワークページャを、24はネットワーク管理部を、221は実メモリ管理部を、222は仮想メモリ管理部を、16はページテーブルを、241は呼設定部を、242は仮想回線管理部を示す。なお、図4では、この実施の形態の説明に必要な部分のみを示しており、そのほかの資源管理(スレッド管理、割り込み管理、デバイス管理)は、従来の仮想記憶方式のOSと同等のものでよく説明に不要なので省略している。

【0047】先に述べたように、実メモリは、「ページ・フレーム」(以下、フレームと略記する)と呼ばれる512バイト〜8Kバイトを単位とした領域に分割されていて、実メモリ14上に存在するデータにアクセスする場合、該当データが含まれているフレームの「フレーム番号」と、そのフレームの先頭から該当データまでの距離を示した「オフセット」を用いてアドレッシングされる。「フレーム番号」と「オフセット」を併せたものを「物理アドレス」と呼ぶ。

【0048】一方、プログラム内では、プログラムごとに一意な「仮想アドレス」と呼ばれるアドレスを用いて

アドレッシングされているため、プログラムの実行時、「仮想アドレス」から「物理アドレス」へ変換を行わなければならない。

【0049】その変換の方法としては、例えば、次のような方法を用いる。仮想アドレスを、上位部分(以下、「仮想フレーム」と呼ぶ)と、下位部分(以下、「オフセット」と呼ぶ)の2つの部分に分割し、下位部分の「オフセット」はそのまま「物理アドレス」の「オフセット」となるようにし、上位部分の「仮想フレーム」をキーとしてページテーブル16を参照して「物理アドレス」の「フレーム番号」に変換する。すなわち、ページテーブル16の仮想フレームに対応するエントリ・ポイントに、予め、フレーム番号を書き込んだ「ページテーブル・エントリ」を書き込んでおき、変換時、仮想フレームに対応するエントリ・ポイントに書かれた「ページテーブル・エントリ」を参照し、このページテーブル・エントリに書かれたフレーム番号を読み出すことによって、仮想フレームに対応する実メモリのフレーム番号を得ることができる。

【0050】また、実メモリ上にデータが存在しない場合は、対応するページテーブル・エントリの「存在ビット」を0とすることによって、仮想アドレスから物理アドレスに変換するために、ページテーブル16をアクセスする際、ページ・フォルトを発生する(プロセッサ15に備わった機能)ことによって、該ページフレームに属するデータ・ブロックを、ハードディスク等から実メモリに転送すると、ページテーブル・エントリに該ページフレームのフレーム番号を登録する処理ルーチンを起動することができる。

【0051】ここで、従来の技術においては、仮想アドレスによって、ATMやEthernet等のネットワークで接続された他の計算機に属する資源にアクセスする場合、該他の計算機との間に設定されたコネクションを使用するために必要な情報は、ページテーブル16には書かれておらず、上記情報を得ようとする場合、種々のテーブルを参照して得なければならず効率が悪かった。

【0052】この実施の形態においては、上記コネクションを使用するのに必要なVPI/VCI値といった仮想回線識別子の情報をページテーブル16に書き込んでおき、上記資源にアクセスするためページテーブル16を参照した際に、上記情報をただちに得られるようにすることによって、高速化を図っている。

【0053】手順としては、詳細は後述するが、仮想フレームに対応するページテーブル・エントリを参照した結果、ページテーブル・エントリの「存在ビット」が1の場合は、仮想フレームに対応するページフレームは実メモリ14中に存在し、ページテーブル・エントリに書かれたフレーム番号を用いて物理アドレスに変換し、実メモリ14をアクセスすることができる。一方、ページ

テーブル・エントリの「存在ビット」が0の場合で、ページテーブル・エントリに該資源を所有する他の計算機間に接続されたコネクションを使用するために必要なVPI/VC I値といった情報が書き込まれている場合、この書き込まれた情報を用いて、上記資源を所有する計算機に対し、仮想アドレスに対応するデータが含まれているデータ・ブロックの転送を要求することができる。

【0054】仮想記憶空間管理部22は、基本的には、従来技術のものと同様である。つまり、仮想アドレス変換部13内のページテーブル16の領域管理と各ページテーブル16内のページテーブル・エントリにパラメータをセットし、仮想アドレス変換部13により仮想アドレスから物理アドレスへの変換ができるように設定を行う。図5は、仮想記憶空間管理部22が扱うページテーブル16の一例である。以下、図5を参照しながら、ページテーブル16の構成例について説明する。

【0055】図5は、「フレーム番号」、「存在ビット」、「処理中ビット」、「ハードディスクビット」、「リモートホストビット」等の情報を格納する場合を示したものである。

【0056】なお、「フレーム番号」と「VPI/VC I値」といった他の計算機間に接続されたコネクションを使用するのに必要な情報と「ハードディスク上のセクター番号」は、同じ場所（図の中では「フレーム番号」と名前をつけた場所）に格納されている。「存在ビット」、「ハードディスクビット」、「リモートホストビット」の値によって、同じ「フレーム番号」の場所に格納された数値の意味が決まる。

【0057】「存在ビット」は、前述したものと同じである。「処理中ビット」は、仮想フレームに対応するページテーブル・エントリに該当するページフレームの準備が終了していない場合に1がセットされる。つまり、ページテーブル・エントリに、VPI/VC I値等の情報が書かれて、そのVPI/VC I値等を使用して、資源を所有する計算機に対し、データの転送を要求し、処理が未完了の場合や、ディスクからデータを転送の処理中（未完了）などの場合である。処理中ビットに1をセットすることによって、他のプログラム等が、同じページテーブル・エントリにアクセスし、再度二重に同じデータの転送を要求することを防止することができる。

【0058】「ハードディスクビット」は、ハードディスク19中のデータが格納されているセクタ番号が、「フレーム番号」欄に書かれていることを示す用途等を使用することができる。

【0059】「リモートホストビット」は、ページテーブル・エントリの「フレーム番号」欄にかかれた数値が、VPI/VC I値等の情報が書かれていることを示し、VPI/VC I値等を使用して、資源を所有する計算機に対し、データの転送を要求する場合などに利用す

る。

【0060】なお、後述するように、データを所有するリモートホストかあるいはリモートホストを検索するためのネーム・サーバ5かを区別するためなどの用途にも使用できる。

【0061】ネットワーク共有メモリ管理部21は、ユーザプログラムなどからの要求により、ローカルホスト（自ホスト）側の仮想記憶空間とリモートホスト側の仮想記憶空間の共有化の設定や管理を行う。

10 【0062】ネットワーク共有メモリ管理部21の機能によって、分散システムを構成することができる。つまり、この機能を利用し、共有領域を設定すると、ユーザやアプリケーションプログラムからは、設定された仮想記憶空間が他のホストと共有され、ここを参照したり、書き込んだりして利用することを通して他のホストの資源を利用できる。

20 【0063】この共有空間はOSが管理して作り出している仮想的なものである。実際には、各ホスト1の上での仮想記憶空間管理部22で実現する。つまり、各ホスト1ごとに、それぞれに仮想記憶空間を別々に管理し保持しているが、ネットワーク共有メモリ管理部21が、仮想記憶空間管理部22の該当する領域のページテーブル・エントリに、予め共有されている旨の設定（仮想回線の識別子の設定など）をしておく。実際にその領域にアクセスがあると、仮想記憶空間管理部22から起動されたネットワークページ23が、他ホストのネットワークページ23と通信を行い、データを転送し、仮想記憶空間の共有を実現する。

30 【0064】ネットワーク管理部24は、ATMの仮想回線の管理を行うところであり、ネットワークインターフェイス制御部12とネットワークインタフェース11を扱う。

【0065】ネットワークページ23とは、各計算機間で共有する領域どうしのデータを実際に転送する機能であり、ネットワーク共有メモリ管理部21により作成されたATMの仮想回線を用いてデータ転送する。クライアント側は、仮想記憶空間管理部22から起動され、サーバ側は、ネットワーク管理部24から起動される。

40 【0066】以上、この実施の形態に係る分散システム全体の概要を説明した。以下では、この分散システムの処理の流れを次の2つに分けて説明する。

(1) 共有領域の対応管理と仮想回線の設定

(2) 共有領域のデータ転送

上記の(1)は、複数の計算機1間で、共有する領域の設定とその共有関係を管理する処理である。概略的には、共有関係にある領域間で、ATMの呼設定を行い、仮想回線を張っておく。その仮想回線の識別子を仮想記憶空間管理部22のページテーブル16に登録しておく、実際のデータ転送を行うときに利用する。

50 【0067】上記の(2)は、共有領域領域に対する参

照や書き込みなどのアクセスが実際に発生し、共有する計算機1どうしでデータの転送が必要になったときに、ネットワークページャ23を用いてデータを転送する処理である。上記の(1)の処理で既に仮想回線が張っているので、それを用いてデータ転送を行う。

【0068】ところで、共有方法は、共有領域の対象が仮想記憶空間そのものか、実記憶空間かによって、図6に示す4種類の共有タイプが考えられる。以下では、この分散システムを、4タイプの共有方法について説明する。

【0069】<タイプ1>タイプ1は、クライアント側の仮想記憶空間の領域とサーバ側の仮想記憶空間の領域を共有領域ごとにネットワーク接続により共有するものである(図7参照)。

【0070】<タイプ2>タイプ2は、クライアント側の仮想記憶空間の領域とサーバ側の実記憶空間の領域を共有領域ごとにネットワーク接続により共有するものである(図14参照)。

【0071】<タイプ3>タイプ3は、クライアント側の実記憶空間の領域とサーバ側の仮想記憶空間の領域を共有領域ごとにネットワーク接続により共有するものである(図15参照)。

【0072】<タイプ4>タイプ4は、クライアント側の実記憶空間の領域とサーバ側の実記憶空間の領域を共有領域ごとにネットワーク接続により共有するものである(図16参照)。

【0073】OSは、共有のタイプに応じて、対応する領域間でATMの仮想回線を張ることになる。

<タイプ1>最初に、図6のタイプ1について説明する。図7に、タイプ1の概要を表す模式図を示す。

【0074】(1)共有領域の対応管理と仮想回線の設定

まず、タイプ1による仮想回線の設定を説明する。図8には、フローチャートの一例を示す。

【0075】タイプ1では、クライアント側の未使用の仮想記憶空間の領域とサーバ側の割当て済みの領域(つまり仮想記憶空間とファイル等の資源の対応が存在する領域)の間を共有化する。

【0076】共有化する仮想記憶空間の領域の指定には、例えば、以下の方式がある。

(i) ユーザプログラムが、明示的に、仮想記憶空間を管理するOSに対して、共有領域の範囲指定とその共有する相手の計算機とその計算機での仮想記憶空間の領域を指定する。つまり、実際のその領域の参照または書き込みなどのアクセスが発生する前に指定する。

【0077】これは、仮想記憶空間にファイルをマッピング(UNIXシステムにおけるmmapシステムコール)して利用する方式における、ファイルシステムの操作におけるOPEN処理に相当する。共有相手の計算機名や相手の領域は、プログラムが直接指定する場合もあ

るし、ネームサーバ5等からの指示であってもよい。

【0078】(ii) OSが、あらかじめ、仮想記憶空間の共有相手の計算機とその相手の計算機上の仮想記憶空間の領域を知っており(起動時に参照する設定ファイルに対応方法が示してありまたはOSのなかに組み込まれており)、共有仮想領域が、最初に参照または書き込みのアクセスをされた時に指定する。上記した(i)と

(ii)は、いずれの場合も共有化処理が起動されるタイミングが異なるだけで処理内容は同じであり、ローカル(クライアント)側のネットワーク共有メモリ管理部21は、共有化する相手の領域を管理するリモート(サーバ)側の計算機1のネットワーク共有メモリ管理部21へ共有化開始の手続きを行う。

【0079】最初に、クライアント側の処理について説明する。図9には、クライアント側のフローチャートの一例を示す。クライアント側のネットワーク共有メモリ管理部21は、上記した(i)または(ii)の方法により、次に示す入力フォーマットのデータをシステムコール等で受取る。

<入力フォーマット>ローカル側仮想記憶空間ID、開始アドレス、サイズ、リモート側ホストID、仮想記憶空間ID、開始アドレス、その他の情報(共有の種類、ユーザID、アクセスの制限、QoS情報など)

そして、ネットワーク管理部24へ、新しいATMの仮想回線をリモート側と設立することを依頼する。

【0080】クライアント側ネットワーク管理部24では、ATMの呼設定セルを作成し、ATM交換機3へ送信する。仮想回線が張られると、その仮想回線に割り当てられたVPI/VCI等の仮想回線の識別子が得られる。

【0081】次に、クライアント側のネットワーク共有メモリ管理部21では、得られた識別子を返り値としてもらい、その値を仮想記憶空間管理部22へ、次に示すフォーマットで、共有空間であることを伝える。

<入力フォーマット>ローカル側仮想記憶空間ID、開始アドレス、サイズ、仮想回線識別子、その他の情報(共有の種類、ユーザID、アクセスの制限、QoS情報など)

仮想記憶空間管理部22では、必要なページテーブルを作成し、ページテーブル・エントリを初期化しパラメータセットする。

【0082】まず、「存在ビット」を0にする。次に、該仮想回線を使用するのに必要なVPI/VCI値といった識別子を、ページテーブル・エントリの「フレーム番号」欄に書き込み、「リモートホストビット」を1にする。

【0083】その他の情報も必要に応じてセットする。最後に、ネットワーク共有メモリ管理部21は、実際のデータ転送時に動作するネットワークページャ23を初期化する。

【0084】以上が、クライアント側の処理である。次に、サーバ側の処理について説明する。図10には、サーバ側の処理の流れを示す。

【0085】サーバ側では、ネットワーク管理部24にて、ATM交換機3からの呼設定セルが受け付けられる。この中の情報にて、共有する仮想記憶空間の領域や相手（クライアント側）の計算機などがわかる。

【0086】次に、この情報はサーバ側のネットワーク共有メモリ管理部21へ送られ、共有化の処理が進められる。ただし、その際に、共有する相手によっては、共有を拒否するようにしてもよい。

【0087】ネットワーク共有メモリ管理部21では、仮想記憶空間管理部22に、共有領域のページテーブル・エントリの内容を確認し、すでにファイルやデータ等の資源が割当て済みの領域であることを確認し、共有に関する情報を必要であればセットする。

【0088】さらに、ネットワークページャ23（サーバ側）が未起動であれば（はじめてこの領域が共有化された場合）、起動し、初期化し、ネットワークページャ23がこの新しい仮想回線を利用してデータを転送できるように諸準備を行う。

【0089】（2）共有領域におけるデータの転送方法次に、共有領域におけるデータ転送について説明する。図11には、処理手順の流れの概要を示す。また、図12と図13は、仮想記憶空間にアクセスしたときの実行時の動作を示したフローチャートの一例である。

【0090】仮想アドレスで示されたデータをアクセス（readまたはwrite）する際、仮想アドレスを物理アドレスに変換するために、ページテーブル16が参照されるが、仮想フレームに対応するページテーブル・エントリの「存在ビット」が1の場合（ステップS1）、ページテーブル・エントリに書かれた「フレーム番号」を使用して、物理アドレスを計算し、計算された物理アドレスを用いて、実メモリ14をアクセスし、求めるデータにアクセスする（ステップS20）。

【0091】「存在ビット」が0、かつ、「処理中ビット」が1の場合（ステップS1、S4）、他のプロセス等が、同じページフレームにアクセスしており、OSで管理されるページャ処理等が、前記ページフレームに対応するデータを、実メモリ14へ転送の処理を行っている途中であることを意味しているため、前記OS等のページャ処理の終了を待って休眠する（ステップS5）。

【0092】前記OS等のページャは、処理を完了すると、該当ページテーブル・エントリの「処理中ビット」を0にし、さらに、処理待ちで休眠したプログラムの実行を再開させる。

【0093】そこで、休眠状態が目覚めると（ステップS6）、前記ページテーブル・エントリに、新たに書き込まれた「フレーム番号」を使用して、物理アドレスを計算し、計算された物理アドレスを用いて、実メモリ1

4をアクセスし、求めるデータにアクセスする（ステップS20）。

【0094】「存在ビット」が0、かつ、「処理中ビット」が0、かつ、「ハードディスクビット」が1の場合（ステップS1、S4、S7）、仮想フレームに対応するページテーブル・エントリの「フレーム番号」の欄は、前記仮想フレームに対応するハードディスク19上のセクタ番号が書かれている。この状態は、たとえば、実メモリ14が不足した場合に、この仮想フレームに対応するページフレームの内容をハードディスク19に一時退避し、このページフレームを明け渡す場合（スワップ・アウトやページ・アウトなどの場合）に発生する。

【0095】この処理手順は、ページャ処理というが、まず、前記データを格納するのに必要なページ・フレームを実メモリ14から確保し、それが、他のプロセス等に使用されないようにロックする（ステップS9）。その値を「フレーム番号」欄に記入し、「ハードディスクビット」を0に、「存在ビット」と「処理中ビット」を1にする（ステップS10）。

【0096】その後、前記セクタ番号を使用して、ハードディスク19から対応するデータブロックを読み出し、前記確保したページ・フレームへとデータ転送を行う。データ転送は、ディスクのデバイスドライバによって行われ、処理の完了を待って、休眠する（ステップS11）。

【0097】デバイスドライバの処理が完了するとOS等から起動されるので、前記ページャ処理を再開する。該当ページテーブル・エントリの「処理中ビット」を0にし、さらに、この処理中ビットのために休眠したプログラムの実行を再開させる（ステップS12）。

【0098】この段階で、実メモリ中にデータブロックがロードされているので（ステップS2）、前記ページテーブル・エントリに、新たに書き込まれた「フレーム番号」を使用して、物理アドレスを計算し、計算された物理アドレスを用いて、実メモリ14をアクセスし、求めるデータにアクセスする（ステップS20）。

【0099】「存在ビット」が0、かつ、「処理中ビット」が0、かつ、「ハードディスクビット」が0、かつ、「リモートビット」が1の場合（ステップS1、S4、S7、S13）、ページャ処理を行なう。仮想フレームに対応するページテーブル・エントリの「フレーム番号」の欄には、前記（1）の共有化処理で設定した、仮想フレームに対応する資源を所有するサーバ計算機への通信に必要なVPI/VCI値といった仮想回線の識別子の情報が書かれているので、前記データを格納するのに必要なページフレームを確保し、それが他のプロセス等に使用されないようにロックする（ステップS15）。その値を「フレーム番号」欄に記入し、「リモートビット」を0に、「存在ビット」と「処理中ビット」を1にする（ステップS16）。

10

20

30

40

50

【0100】その後、前記VPI/VCI値といった仮想回線の識別子の情報を用いて、クライアント側（ローカル側）のネットワークページ23を起動し、データ転送を依頼する（ステップS17）。処理の完了を待つて、休眠する（ステップS18）。

【0101】クライアント側ネットワークページ23では、指示された仮想回線識別子を用いて、ネットワーク管理部24を通じて、リモート側の計算機のネットワークページ23に対して、前記データ・ブロックの転送を要求するセルを作成し送信する。そのセルには、以下の内容が含まれる。

【0102】開始アドレス、サイズ、データ転送の方向（クライアント側→サーバ側、サーバ側→クライアント側）、その他の情報（共有の種類、ユーザIDなど）その後、上記要求セルに回答したデータセルが送られてくるのでその内容を対応するページフレーム内に書き込む。

【0103】前記ページ処理は、上記ネットワークページの処理が完了し、処理が再開されると、該当ページテーブル・エントリの「処理中ビット」を0にし、さらに、この処理中ビットのために休眠したプログラムの実行を再開させる（ステップS19）。

【0104】この段階で、実メモリ中にデータブロックがロードされているので（ステップS2）、前記ページテーブル・エントリに、新たに書き込まれた「フレーム番号」を使用して、物理アドレスを計算し、計算された物理アドレスを用いて、実メモリ14をアクセスし、求めるデータにアクセスする（ステップS20）。

【0105】なお、「存在ビット」が0、かつ、「処理中ビット」が0、かつ、「リモートビット」が0の場合（ステップS1、S4、S7）、エラー処理を行う（ステップS21）。

【0106】リモート側の計算機のネットワーク管理部24では、上述のデータ転送要求セルを受け、その仮想回線の識別子から、この仮想回線を受け持つサーバ側のネットワークページ23を特定し、セルの内容を伝える。

【0107】セルの内容が伝わったサーバ側のネットワークページ23は、上記を転送要求セルを受け、指示された、仮想記憶空間の開始アドレスからサイズ分をデータブロックを読みとり、その内容をセルに入れてクライアント側へ同じ仮想回線を使って送る。

【0108】なお、データ転送には、AAL5などの適切なトランスポートプロトコルで、処理されて送られるものとする。これは、データの内容の順序やデータの抜け、データの内容のエラー、データの再送などを行い、両ネットワークページ23間でのデータ完全性を保証するものである。なお、ATMのネットワーク側で既にある程度の処理がされているため、計算機側での処理は軽い。

【0109】上記の実行は、クライアント側で共有された仮想記憶空間のデータを参照した場合の動作であるが、逆に、クライアント側で書き込みが発生した場合、ネットワークページ23間のデータ転送の方向は逆方向になる。

【0110】ここで、以上説明した主な処理をクライアント側とサーバ側の各部ごとにまとめておく。

ネットワーク共有メモリ管理部21（クライアント側）

10 共有化設定要求を受け付ける。

【0111】ネットワーク管理部24へ、共有相手の計算機との間での仮想回線の呼設定を依頼する。仮想記憶空間管理部22で管理しているページテーブルの該当するエントリに仮想回線識別子等をセットする。

【0112】 ネットワーク管理部24（クライアント側）

ATM交換機3との間で呼設定を行い、新しい仮想回線を設立する。ATMの呼設定時の付加情報機能（呼設定セル中のusr-usr情報、または、呼設定時にend-endホスト間で交換できるusr-usrメッセージセルなど）を利用して、この仮想回線の利用目的等を相手の計算機へ接続要求時に伝える。この仮想回線は、この共有化領域のデータ転送に使うための専用である。

【0113】仮想回線のフローコントロール、エラー処理、ふくそう処理などを行う。

ネットワークページ23（クライアント側）

仮想記憶空間管理部22から、実際にデータ転送の要求があるときに呼ばれる。

30 【0114】仮想記憶空間へのアクセスの種類（readまたはwrite）とそのアドレスから必要な大きさの実メモリを確保し、転送に必要な仮想アドレスの開始位置とサイズ等を相手計算機に要求する。

【0115】readの場合は、相手計算機からデータが送られてくるので、それを実メモリ14に格納する。writeの場合は、転送要求セルを送信した後、データをセルに入れて送信する。

【0116】 ネットワーク管理部24（サーバ側）

40 ネットワーク管理部24では、ATM交換機3からの呼設定要求セルを受けつけ、新しい仮想回線を設立する。

【0117】呼設定時の付加情報によって、新しい回線の利用目的を知り、ネットワーク共有のための新規仮想回線の場合、ネットワーク共有メモリ管理部21（サーバ側）に制御を移す。

【0118】以後、転送要求セルがくると、この仮想回線を担当するネットワークページ23に制御を移す。仮想回線のフローコントロール、エラー処理、ふくそう処理などを行う。

50 【0119】 ネットワーク共有メモリ管理部21（サーバ側）

新規の仮想記憶空間共有依頼が、ネットワーク管理部 24 からくる。新仮想回線を利用して、相手側（クライアント側）のネットワーク共有メモリ管理部 21 と通信し初期設定を行う。つまり、共有化するサーバ側の仮想記憶空間 ID と仮想アドレスの開始位置、サイズ等が伝えられる。これを仮想記憶空間管理部 22 に伝え、共有可能かどうか検査する。

【0120】この共有領域を担当するネットワークページ 23 を起動し、得られた仮想回線の識別子を教える。以後、仮想回線から read 要求あるいは write 10 要求があるごとに、その対応する仮想アドレスの領域に対してデータを読み書きする。

【0121】 ネットワークページ 23（サーバ側）ネットワーク管理部 24 から、実際にデータ転送の要求があるときに呼ばれる。

【0122】対応する仮想記憶空間の領域のデータ転送方向（read または write）と、データ転送する仮想アドレスの開始位置とサイズ等が相手計算機から伝えられる。

【0123】 read の場合は、相手計算機へ仮想記憶空間からデータブロックを読みとり、それをセルに入れて送信する。write の場合は、転送要求セルを受信し、そのセルで指示されている、開始アドレスからサイズ分の大きさの領域に、以後送られてくるセル内のデータを該当領域に書き込む。

【0124】この実施の形態によれば、複数の計算機を ATM ネットワークを介して結合した分散システムにおいて、仮想記憶空間または実記憶空間（実メモリ）の一部を共有化し、共有領域ごとに専用に設定しておいた ATM の仮想回線を利用して直接、高速なデータ転送を行なうことができるので、計算機側には複雑で効率の悪い通信プロトコル処理が必要とせず、効率的な記憶空間の共有が実現できる。つまり、その共有空間に対して参照や書き込み（read/write）が発生しても、その都度、呼設定不要で、ただちにデータ転送ができる。

【0125】また、呼設定、エラー検出、ふくそう制御などデータ転送中の制御は ATM 交換機側で大部分対応してくれるので、計算機側はその機能に不足する部分（例えばデータエラーにおける訂正や再送）だけ管理すればいいので、計算機側の処理が大幅に軽減する。

【0126】さらに、ATM 交換機を利用することによって、ある程度の距離が離れた計算機どうしを結んで分散システムにできるので、構成の柔軟性が良い。また、他の計算機に属する資源にアクセスする場合、仮想アドレスを物理アドレスに変換する際に用いるページテーブルを参照するだけで、他の計算機へのコネクションを使用するために必要な仮想回線識別子（VPI/VCI 値等）を得ることができる。

【0127】このため、他の計算機に属する資源のうち、必要な大きさのデータブロックの転送を要求する場

合、従来のように種々のテーブルを参照することなく、ページテーブルから得られた仮想回線識別子を用いて要求セルを作成することができ、効率良く要求セルを作成することができる。

【0128】なお、単一仮想記憶空間方式の OS の場合、仮想記憶空間は各計算機に一つしか存在しないので、各計算機の仮想記憶空間すべてを共有化し、どの計算機からも同じように仮想記憶空間を利用できる構成にすることもできる。この場合、全仮想記憶空間を 1 つの共有領域とするのではなく、仮想記憶空間を複数の共有領域に分割し、それぞれ別々に共有化処理を行っても良い。

【0129】<タイプ 2>次に、図 6 のタイプ 2 について説明する。上記したタイプ 1 では、共有する領域が各計算機の仮想記憶空間どうしであったが、このタイプ 2 では、クライアント側の仮想記憶空間の領域とサーバ側の実記憶空間（実メモリ）の領域の間を共有化する。

【0130】図 14 に、タイプ 2 の概要を表す模式図を示す。図中、45 はクライアント側計算機に備え付けられたハードディスクを示す。このタイプ 2 は、サーバ側にデータブロックが存在するが、まだ、サーバ側のプログラム等がそのデータブロックを利用しておらず、そのデータブロックを仮想記憶空間に張り付けて利用していない状況での利用に適する。あるいは、例えば、サーバ側では、利用プログラムは存在せず、データブロックを管理するだけであるような構成での利用に適する。つまり、共有空間の利用は、サーバ側からは行われず、各クライアント側からだけの利用を想定する。この場合、データブロックは、サーバ内の磁気ディスク 45 等に永続ファイル等として格納されている。

【0131】共有領域の仮想回線の設定は、タイプ 1 とほぼ同様である。ただし、呼設定時に、サーバ側の共有領域を指定するのに、タイプ 1 における「仮想記憶空間 ID、開始アドレス、サイズ」に代わり、「実メモリの開始アドレス、サイズ」によることになる。しかし、サーバ側では、UNIX 等の汎用 OS における実メモリの管理方法として、通常、実メモリは仮想記憶空間に一時的に割当て、ディスクの内容（データブロック）を一時的に蓄えるバッファとして利用する。したがって、このような OS の場合、呼設定時に実メモリのアドレスを直接明示して指定しても実際的ではない。したがって、データブロック（つまり永続ファイル）をアドレスの代わりにファイル名などで直接指定し、サーバ側では、ネームサーバやファイルシステムにより、指定されたファイル名からデータブロックが格納されている磁気ディスクの位置を求める。そして、必要な大きさの実メモリ 14 を確保し、その内容を読みだして実メモリ 14 へ転送する。この実メモリ 14 の領域を共有領域としてサーバ側のネットワークページ 23 が管理する。

【0132】クライアント側の共有領域の読み書きに伴

うデータ転送はタイプ1と同様である。このタイプ2では、サーバ側で仮想記憶空間に実メモリを対応され、データを読み込むページフォルト処理が不要である。また、予め、サーバ側の実メモリにディスク等からデータが転送されているので、クライアント側への仮想回線を介するデータ転送が高速に行える。

【0133】<タイプ3>次に、図6のタイプ3について説明する。上記したタイプ1では、共有する領域が各計算機の仮想記憶空間どうしであったが、このタイプ3では、クライアント側の実記憶空間の領域とサーバ側の仮想記憶空間の領域の間を共有化する。そして、共有化開始処理を完了したときにまたは共有化した領域をプログラム処理実体がアクセスする前に、あらかじめデータ転送を行なっておくものである。

【0134】図15に、タイプ3の概要を表す模式図を示す。

(1) 共有領域の仮想回線設定

仮想回線の設定方法は、タイプ1と類似する方法になる。

【0135】クライアント側のネットワーク共有メモリ管理部21は、共有化する相手のリモートの領域を管理する計算機のネットワーク共有メモリ管理部21へ共有化開始の手続きを行う。タイプ1では、クライアント側の共有領域の指定方法として、仮想記憶空間の一部の領域を指定した。

【0136】この実施の形態でも同様に、クライアント側の仮想記憶空間の領域を指定しても良い。この場合は、タイプ1と同様な入力フォーマットで指定する。

<入力フォーマット>ローカル側仮想記憶空間ID、開始アドレス、サイズ、リモート側ホストID、仮想記憶空間ID、開始アドレス、その他の情報（共有の種類、ユーザID、アクセスの制限、QoS情報など）

この場合は、クライアント側の仮想記憶空間管理部22で、指定された仮想記憶空間の領域と同じサイズの実メモリを確保し、共有領域用として利用する。

【0137】他の方法として、クライアント側の仮想記憶空間とは無関係に、実メモリを確保しても良い。この場合、予め、共有領域に必要な大きさの実メモリを直接確保する。この方法では、クライアント側のOS等が提供するシステムコール等を利用し、仮想記憶空間管理部22内の実メモリ管理部221に対して要求する。

【0138】共有指定は、次のフォーマットで行う。
<入力フォーマット>ローカル側実メモリ開始アドレス、サイズ、リモート側ホストID、仮想記憶空間ID、開始アドレス、その他の情報（共有の種類、ユーザID、アクセスの制限、QoS情報など）

この場合、クライアント側のユーザプログラム等は、実メモリのアドレスを直に指定してデータを利用する（OSが提供する実メモリのアドレスを指定してデータをアクセスするシステムコール等を利用する）。または、任

意の仮想記憶空間に実メモリをマップして利用する。

【0139】クライアント側のネットワーク共有メモリ管理部21は、先のフォーマットのデータをシステムコール等で受取る。そして、ネットワーク管理部24へ新しいATMの仮想回線をリモート側と設立することを依頼する。

【0140】クライアント側のネットワーク管理部24では、ATM交換機3へATMの呼設定セルを作成し、ATM交換機3へ送信する。仮想回線が張られると、その仮想回線に割り当てられたVPI/VCI等の仮想回線の識別子が得られる。

【0141】次に、クライアント側のネットワーク共有メモリ管理部21では、実際のデータ転送時に動作するネットワークページ23を初期化する。ネットワーク管理部24で得られた識別子を返り値としてもらい、その値をネットワークページ23に伝えて、データ転送を開始を依頼する。

【0142】以上で、クライアント側の共有化開始時の処理が終わる。次に、サーバ（リモート）側の処理を説明する。サーバ側では、ネットワーク管理部24にて、ATM交換機3からの呼設定セルが受け付けられる。この中の情報にて、共有する仮想記憶空間の領域や相手（クライアント）の計算機などがわかる。

【0143】次に、この情報をサーバ側のネットワーク共有メモリ管理部21へ送られ、共有化の処理が進められる。この際、共有する相手によっては、拒否してもよい。ネットワーク共有メモリ管理部21では、仮想記憶空間管理部22に、共有領域のページエントリの内容を確認し、すでにファイル等の資源が割当済みの領域であることを確認し、共有に関する情報を必要であればセットする。さらに、ネットワークページ23（サーバ側）を起動し、このネットワークページ23がこの新しい仮想回線を利用してデータを転送できるように諸準備を行う。

【0144】（2）共有領域におけるデータの転送方法次に、共有領域におけるデータ転送について説明する。タイプ1と異なり、クライアント側の共有領域には、すでに実メモリが割り当てられているので、その領域をプログラム等の処理実体がアクセスしても、ページフォルトは発生しない。この実施の形態の場合は、クライアント側での実際のアクセスが発生した時にデータ転送するのではなく、上記仮想回線の設定時に、引き続いてデータ転送を行う（すなわちクライアント側のアクセスとは無関係である）。

【0145】まず、設定時にサーバ側からデータをクライアントに転送しておき、クライアント側の参照アクセスの準備を完了させる。その後、クライアント側の書き込みがあった場合は、別手段（システムコールで指示したり、定期的に行ったり、共有領域の解消時に実行したりするなど）によりサーバ側へデータ転送を行う。

【0146】クライアント側ネットワークページ23では、指示された仮想回線識別子を用いて、ネットワーク管理部24を通じて、リモート側の計算機のネットワークページ23に対して、前記データ・ブロックの転送を要求するセルを作成し送信する。

【0147】そのセルには、以下の内容が含まれる。開始アドレス、サイズ、データ転送の方向（クライアント側→サーバ側、サーバ側→クライアント側）、そのほかの情報（共有の種類、ユーザIDなど）

その後、上記要求セルに回答したデータセルが送られてくるのでその内容を対応する実メモリ（ページフレーム）に書き込む。書き込みが完了すると以後の処理は、別手段（システムコールで指示したり、定期的に行ったり、共有領域の解消時に実行したりするなど）からの指示によって、サーバ側との間で転送を行う。

【0148】リモート側の計算機のネットワーク管理部24では、上述のデータ転送要求セル受け、その仮想回線の識別子から、この仮想回線を受け持つサーバ側のネットワークページ23を特定し、セルの内容を伝える。

【0149】指示されたサーバ側のネットワークページ23は、上記を転送要求セルを受け、指示された仮想記憶空間の開始アドレスから該当サイズ分のデータブロックを読みとり、その内容をセルに入れてクライアント側へ同じ仮想回線を使って送る。

【0150】タイプ3では、複数のページフレーム（実メモリ）を、予めクライアント側の1つの共有領域に設定できる。クライアント側では、共有化領域である実メモリを仮想記憶空間にマッピングして利用すると、ユーザプログラム等が仮想記憶空間にアクセスした時には、既に実メモリが割り当てられているので、ページフォルトが発生せず、ページフォルト処理が不要になり、即座にデータをアクセスできる。したがって、高速なユーザプログラムの実行が可能となる。

【0151】なお、データ転送には、AAL5などの適切なトランスポートプロトコルで、処理されて送られるものとする。これは、データの内容の順序やデータの抜け、データの内容のエラー、データの再送などを行い、両ネットワークページ23間でのデータ完全性を保証するものである。

【0152】ATMのネットワーク側で既にある程度の処理がされているため、計算機側での処理は軽い。ここで、以上説明した主な処理をクライアント側とサーバ側の各部ごとにまとめておく。

【0153】ネットワーク共有メモリ管理部21（クライアント側）

共有化設定要求を受け付ける。ネットワーク管理部24へ、共有相手の計算機との間での仮想回線の呼設定を依頼する。

【0154】指示された実メモリを確保し、または、必

要な大きさの実メモリを確保し、ネットワークページ23にデータ転送を依頼する。

ネットワーク管理部24（クライアント側）

ATM交換機3との間で呼設定を行い、新しい仮想回線を設立する。

【0155】ATMの呼設定時の付加情報機能（呼設定セル中のusr-usr情報、または、呼設定時にend-endホスト間で交換できるusr-usrメッセージセルなど）を利用して、この仮想回線の利用目的他を相手のマシンへ接続要求時に伝える。この仮想回線は、この共有化領域のデータ転送に使うための専用である。

【0156】仮想回線のフローコントロール、エラー処理、ふくそう処理などを行う。

ネットワークページ23（クライアント側）

ネットワーク共有メモリ管理部21から、共有領域設定完了後、呼ばれる。初期化時に知らされている仮想回線を利用して、転送に必要なアドレスの開始位置とサイズ等を相手計算機に要求する。

【0157】readの場合は、相手計算機からデータが送られてくるので、それを実メモリ14に格納する。writeの場合は、転送要求セルを送信した後、実メモリ14のデータをセルに入れて送信する。

【0158】ネットワーク管理部24（サーバ側）

ネットワーク管理部24では、ATM交換機3からの呼設定要求セルを受けつけ、新しい仮想回線を設立する。

【0159】呼設定時の付加情報によって、新しい回線の利用目的を知り、ネットワーク共有のための新規仮想回線の場合、ネットワーク共有メモリ管理部21（サーバ側）に制御を移す。

【0160】以後、転送要求セルがくると、この仮想回線を担当するネットワークページ23に制御を移す。仮想回線のフローコントロール、エラー処理、ふくそう処理などを行う。

【0161】ネットワーク共有メモリ管理部21（サーバ側）

新規の仮想記憶空間共有化依頼が、ネットワーク管理部24からくる。新仮想回線を利用して、相手側（クライアント側）のネットワーク共有メモリ管理部21と通信し初期設定を行う。つまり、共有化するサーバ側の仮想記憶空間IDと仮想アドレスの開始位置、サイズ等が伝えられる。これを仮想記憶空間管理部22へ伝え、共有可能かどうか検査する。

【0162】この共有領域を担当するネットワークページ23を起動し、得られた仮想回線の識別子を教える。以後、仮想回線から、read要求、write要求があるごとにその仮想アドレスに対応する領域に対してデータを読み書きする。

【0163】ネットワークページ23（サーバ側）

ネットワーク管理部24から、実際にデータ転送の要求

があるときに呼ばれる。

【0164】対応する仮想記憶空間の領域のデータ転送方向（readまたはwrite）と、データ転送する仮想アドレスの開始位置とサイズが相手計算機から伝えられる。

【0165】readの場合は、相手計算機へ仮想記憶空間からデータブロックを読みとり、それをセルにいでて送信する。writeの場合は、転送要求セルを受信し、そのセルで指示されている、開始アドレスからサイズ分の大きさの領域に、以後送られてくるセル内のデータを該当領域に書き込む。

【0166】＜タイプ4＞次に、図6のタイプ4について説明する。このタイプ4では、クライアント側の実記憶空間の領域とサーバ側の実記憶空間の領域の間を共有化する。そして、タイプ3と同様に共有化開始処理を完了したときにまたは共有化した領域をプログラム処理実体がアクセスする以前に、データ転送を行なっておくものである。

【0167】図16に、タイプ4の概要を表す模式図を示す。タイプ3におけるサーバ側の共有化対象がタイプ2におけるサーバ側と同様に実メモリに変わったものである。クライアント側の処理はタイプ3と同様であり、サーバ側の処理はタイプ2と同様である。

【0168】＜タイプ4の応用＞次に、上記のタイプ4に加え、クライアント側とサーバ側で共有領域のデータ転送の同期を取る実施の形態を説明する。

【0169】図17と図18には、その一例として、サーバ側からクライアント側へ動画データを送送し、クライアント側で動画を表示する場合を示している。図中、78はハードディスクである。

【0170】共有領域の大きさは、仮想記憶空間の大きさより実メモリの大きさが大きい。実メモリを複数の領域に分割して管理する。この場合、クライアント側とサーバ側でそれぞれの仮想記憶空間に複数の実メモリ領域を繰り返し割当てて。

【0171】この実施の形態では、割り当てられた実メモリの大きさが仮想記憶空間の大きさの2倍であるとする。したがって、実メモリを2つの領域に分割して利用するものとする。

【0172】クライアント側では、画像表示プログラムが動作し、サーバ側では、画像データ読み込みプログラムが動作している。クライアント側では、仮想記憶空間にマッピングされている実メモリ（ページフレームA）には、既に1つの動画データが入っているのを、それを読みだして表示する。

【0173】その間に、クライアント側では、仮想記憶空間に張られていない別の実メモリ（ページフレームB）が仮想回線を介してサーバ側からデータ転送される。転送が終わり、かつ、画像表示プログラムによって、現在仮想記憶空間にぱッピングしている実メモリか

らの画像表示が終わると、画像表示プログラムは、ネットワークページ23に対して、同一仮想記憶空間に割り当てられる実メモリの切り替え要求を出す（システムコールなどで指示する）。

【0174】すると、ネットワークページ23では、サーバからのデータ転送が完了していると、その別の実メモリ（ページフレームB）を、現在張られている実メモリ（ページフレームA）に代わって張るように仮想記憶空間管理部22へ依頼する。

10 【0175】以上の処理が終わると、クライアント側の画像表示プログラムでは、新しい画像データの入った実メモリを同じ仮想記憶空間の領域から読みだして表示することができる。

【0176】クライアント側のネットワークページ23では、実メモリに余裕がある間サーバ側にデータ転送を要求する。そして、送られてきたデータを空き実メモリ（ページフレームA）に転送する。

【0177】サーバ側のネットワークページ23も、クライアント側と同様に動作する。つまり、空の実メモリ（ページフレームC）が仮想記憶空間に張られている。その仮想記憶空間に対して、サーバ側の画像読み込みプログラムは、ディスクに蓄えられている画像データを順次読み込み、（必要に応じて画像データの加工を行いながら）その領域へ書き込む。領域分の書き込みが終了したら、ネットワークページ23に対して、同一仮想記憶空間に割り当てられる実メモリの切り替え要求を出す（システムコールなどで指示する）。

【0178】すると、ネットワークページ23では、クライアントへのデータ転送が完了し、かつ、空いた実メモリ（ページフレームD）が存在していると、現在張られている実メモリ（ページフレームC）に代わって張るように仮想記憶空間管理部22へ依頼する。

【0179】この処理が終わると、サーバ側の画像読みだしプログラムでは、新しい画像データをディスクから読みだし、同じ仮想記憶空間の領域に書き込む。サーバ側のネットワークページ23では、データ転送が終わっていない実メモリ（ページフレームC）が存在する限り、クライアントへデータ転送を行い、空の実メモリを補充する。

40 【0180】このようにすれば、計算機間で高速にデータを転送することができる。

《ネームサーバについて》ところで、上記した実施の形態における仮想回線の設定は、共有領域どうしの専用の仮想回線が設定されていない場合は、前記仮想回線を設定するように呼設定セルを送信し、仮想回線が設定された後、前記仮想回線を使用するのに必要なVPI/VCI値といった情報を、ページテーブル・エントリに書き込むものであった。

【0181】しかし、前記資源を所有する計算機の所在がユーザやプログラムから伝わらず、わからない場合

(例えば、分散ファイルシステムにおけるファイル名をユーザが指定して共有領域とした場合；この場合、ファイル名は位置透過性であり、実際のファイル資源を所有する計算機の所在はわからない) は、前記所在を知るために、分散システム全域についてデータ・ブロック等の資源の所在を管理しているネームサーバ等々に問い合わせ、その結果知り得た計算機に対し、必要に応じてコネクションを設定し、前記コネクションを使用するのに必要な V P I / V C I 値といった情報を、ページテーブル・エントリに書き込んでも良い。

【0182】また、予め、自計算機が所有しない資源に関するページテーブル・エントリには、直接ネームサーバとの接続のために仮想回線の識別子をページテーブル16にデフォルト値として記入しておいても良い。この場合、資源を所有する計算機(オーナーを計算機)を知る必要が生じた際に、図19のフローチャートに示すように、オーナー計算機を知るために、該当するページテーブル・エントリに書き込まれている V P I / V C I 値等の仮想回線識別子を用いて、ネームサーバに問い合わせを行う(ステップS31)。ネームサーバから、オーナー計算機や共有化領域の範囲など領域の共有化処理に必要な情報が返る(ステップS32)。この情報を利用して、オーナー計算機との間に共有領域の設定をするように、ネットワーク共有メモリ管理部21に要求する(ステップS33)。

【0183】以降、共有領域の設定処理は同様で、ネットワーク共有メモリ管理部21では、上記オーナー計算機の共有領域に対する仮想回線を設定し、その仮想回線の V P I / V C I 等の識別子を、共有化範囲の該当するページテーブル・エントリに書き込む。その他、ネットワークページの初期化作業も同様に行なう。

【0184】この場合、データ転送時のページャ処理を若干変更する。すなわち、クライアントでページフォルトが発生し、ページテーブル・エントリの「存在ビット」が0、かつ、「処理中ビット」が0、かつ、「ハードディスクビット」が0、かつ、「リモートビット」が1の場合(図12、13のステップS1、S4、S7、S13)、仮想フレームに対応するページテーブル・エントリの「フレーム番号」の欄には、その共有領域のデータ転送を行なう仮想回線の識別子が書かれているのではなく、その共有領域の所在を管理するネーム・サーバ等への通信に必要な V P I / V C I 値といった情報(この値は、データ転送用の V P I / V C I 値とは異なる特別な値であるので、それと判明する)が書き込まれているので、前記 V P I / V C I 値といった情報を用いて、前記共有領域のオーナー計算機や共有領域の範囲などを問い合わせるセルを作成/送信し、前記ネームサーバ等からの返答を待つ。

【0185】前記ネーム・サーバからの返答を受信した場合、前記返答中に書かれている前記資源を所有する

計算機への通信に必要な情報から、共有領域の設定処理を行う(ステップS31～S33)。

【0186】それ以降の処理は、図12、13のステップS14からの処理と同様である。本発明は、上述した実施の形態に限定されるものではなく、その技術的範囲において種々変形して実施することができる。

【0187】

【発明の効果】本発明によれば、複数の計算機をネットワーク(例えばATMネットワーク)を介して結合した分散システムにおいて、仮想記憶空間または実記憶空間の全部または一部を共有化し、共有領域ごとに専用に設定しておいた仮想回線を利用して直接、高速なデータ転送を行なうことができるので、計算機側には複雑で効率の悪い通信プロトコル処理が必要とせず、効率的な記憶空間の共有が実現できる。つまり、その共有空間に対して参照や書き込み(read/write)が発生しても、その都度、呼設定不要で、直にデータ転送ができる。

【0188】また、呼設定、フロー制御、ふくそう制御、エラー検出、QoS管理などデータ転送中の制御はネットワーク側(例えばATM交換機側)で対応してくれるので、計算機側はその機能に不足する部分(例えばデータエラーにおける訂正や再送)だけ管理すればいいので、計算機側の処理が大幅に軽減する。

【0189】さらに、ATMネットワークのように高速なネットワークを利用することによって、ある程度の距離が離れた計算機どうしを結んで分散システムにできるので、構成の柔軟性が良い。

【0190】また、本発明によれば、他の計算機に属する資源にアクセスする場合、仮想アドレスを物理アドレスに変換する際に用いるページテーブルを参照するだけで、他の計算機へのコネクションを使用するために必要な仮想回線識別子(V P I / V C I 値等)を得ることができる。

【0191】このため、他の計算機に属する資源のうち、必要な大きさのデータブロックの転送を要求する場合、従来のように種々のテーブルを参照することなく、ページテーブルから得られた仮想回線識別子を用いてデータ転送の要求セルを作成することができ、効率良く要求セルを作成することができる。

【0192】また、本発明によれば、前記ページテーブルの前記仮想回線識別子を記入する項目にデフォルト値として、データとこれを所有する計算機(オーナー計算機)との対応を管理するネームサーバに接続するのに用いる仮想回線識別子を記入しておけば、データを所有する計算機(オーナー計算機)を知る必要が生じた際に、該当するページテーブル・エントリに書き込まれている仮想回線識別子を用いて、速やかにネームサーバに問い合わせを行なうことができる。

【0193】また、本発明によれば、クライアント側の

実記憶空間の全部または一部の領域と、サーバ計算機の仮想記憶空間または実記憶空間の全部または一部の領域とを共有化する場合、クライアント側の共有領域には、すでに実記憶空間（実メモリ）が割り当てられているので、ページフォルトは発生しない。この場合、クライアント側での実際のアクセスが発生した時にデータ転送するのではなく、仮想回線の設定時に、引き続いてデータ転送を行う（すなわちクライアント側のアクセスとは無関係である。）

したがって、ユーザプログラムやスレッド等の処理実体が仮想記憶空間にアクセスしたときには、既に実記憶空間（実メモリ）が割り当てられているので、ページフォルトが発生せず、ページフォルト処理が不要になり、即座にデータをアクセスできる。したがって、例えば、リアルタイム処理等の高速なユーザプログラムの実行が可能となる。

【0194】また、本発明によれば、クライアント計算機の仮想記憶空間の一部の領域と、サーバ計算機の仮想記憶空間の一部の領域とを共有化するとともに、共有する仮想記憶空間の1つの領域には複数の実メモリの領域を割当て、仮想記憶空間の領域と現在対応していない

（マッピングしていない）実記憶空間間で専用の仮想回線を用いてデータ転送を行うとともに、データ転送の完了した実メモリの領域を仮想記憶空間の領域に対応させる（マッピングする）ように、実メモリの領域を順次切り替えるので、計算機間で仮想記憶空間の領域の大きさ単位に連続して高速にデータを転送することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態に係る分散システム全体を示す基本構成図

【図2】この実施の形態に係る計算機内部の要部構成を示す基本構成図

【図3】この実施の形態に係る計算機内部の物理構成の一例を示す物理構成図

【図4】この実施の形態に係るOSの要部構成を示す図

【図5】この実施の形態に係るページテーブルの一例を示す図

【図6】記憶空間の共有方法のバリエーションを表す図

【図7】クライアント側の仮想記憶空間の領域とサーバ側の仮想記憶空間の領域を共有するタイプ1の概要を表す模式図

【図8】共有領域の対応管理と仮想回線の設定の手順を

示す図

【図9】共有領域設定時のクライアント側の動作を示すフローチャート

【図10】共有領域設定時のサーバ側の動作を示すフローチャート

【図11】共有領域におけるデータの転送の手順を示す図

【図12】仮想記憶空間にアクセスしたときの動作を示すフローチャート

10 【図13】仮想記憶空間にアクセスしたときの動作を示すフローチャート

【図14】クライアント側の仮想記憶空間の領域とサーバ側の物理空間の領域を共有するタイプ2の概要を表す模式図

【図15】クライアント側の物理空間の領域とサーバ側の仮想記憶空間の領域を共有するタイプ3の概要を表す模式図

【図16】クライアント側の物理空間の領域とサーバ側の物理空間の領域を共有するタイプ4の概要を表す模式図

【図17】タイプ4に加えクライアント側とサーバ側で共有領域のデータ転送の同期を取る場合を説明するための図

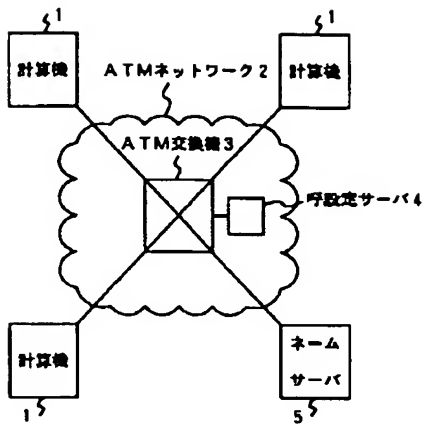
【図18】タイプ4に加えクライアント側とサーバ側で共有領域のデータ転送の同期を取る場合を説明するための図

【図19】資源を所有する計算機が分からない場合の処理の流れを示すフローチャート

【符号の説明】

30 1…計算機、2…ATMネットワーク、3…ATM交換機、4…呼設定サーバ、5…ネームサーバ、11…ネットワークインタフェース、12…ネットワークインタフェース制御部、13…仮想アドレス変換部、14…実メモリ、15…プロセッサ、16…ページテーブル、17…システムバス、18…ハードディスク制御部、19…ハードディスク、21…ネットワーク共有メモリ管理部、22…仮想記憶空間管理部、23…ネットワークページャ、24…ネットワーク管理部、151…バスインタフェース部、152…エグゼキューション部、153…TLB、221…実メモリ管理部、222…仮想メモリ管理部、241…呼設定部、242…仮想回線管理部

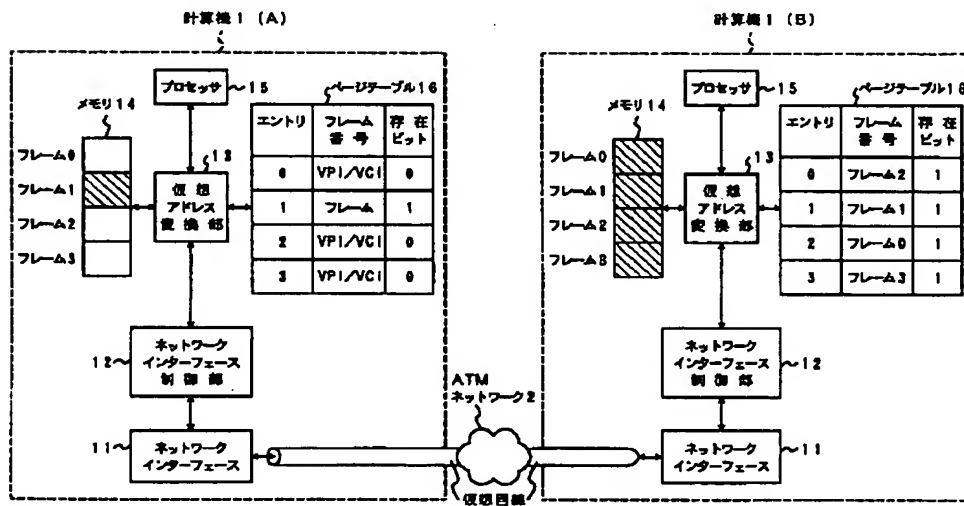
【図 1】



【図 5】

エントリ	フレーム番号	存在 ビット	処理中 ビット	ハード ディスク	リモート ホスト	スレッド ID	その外 の情報 (QoS 情報など)
0	VP1/VC1	0	0	0	1	a	1
1	フレーム1	1	0	0	0	a	0
2	セクタ4	0	1	1	0	b	0
3	VP1/VC1	0	0	0	0	b	0

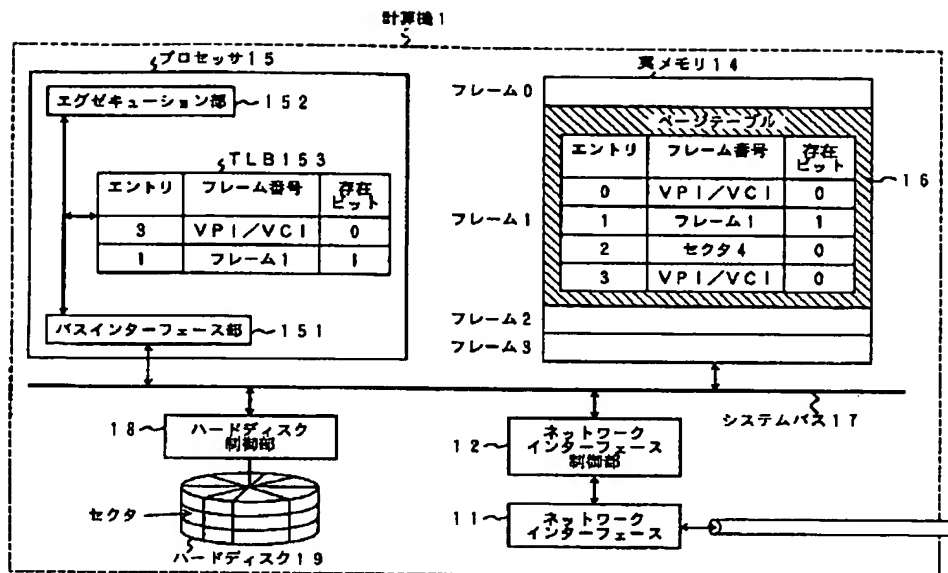
【図 2】



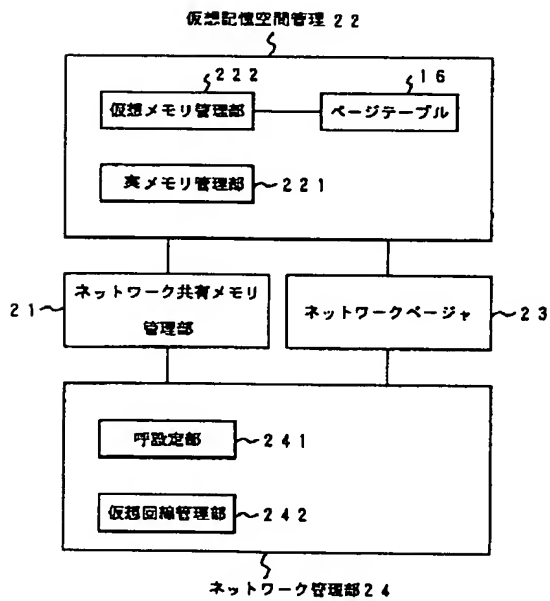
【図 6】

	クライアント側	サーバ側
タイプ1	仮想記憶空間	仮想記憶空間
タイプ2	仮想記憶空間	実記憶空間
タイプ3	実記憶空間	仮想記憶空間
タイプ4	実記憶空間	実記憶空間

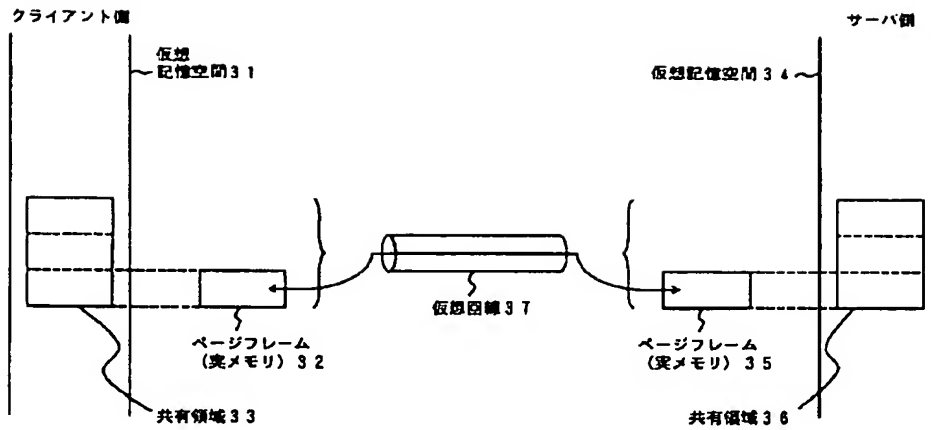
【図3】



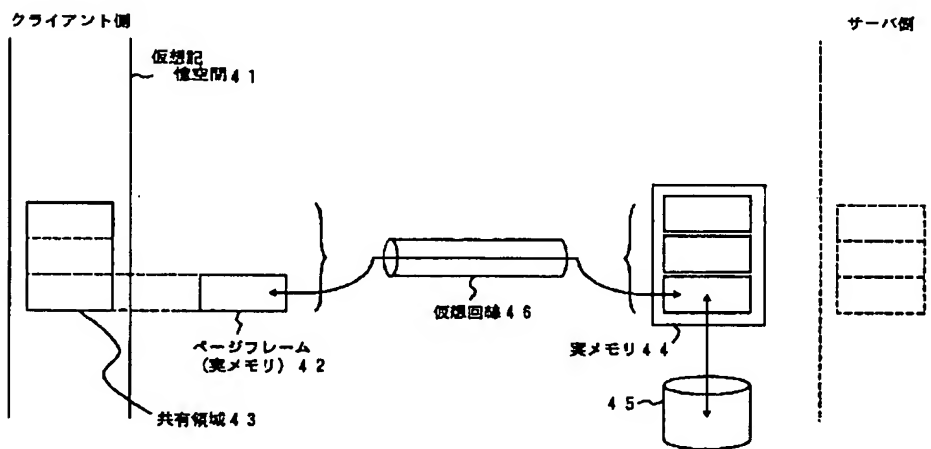
【図4】



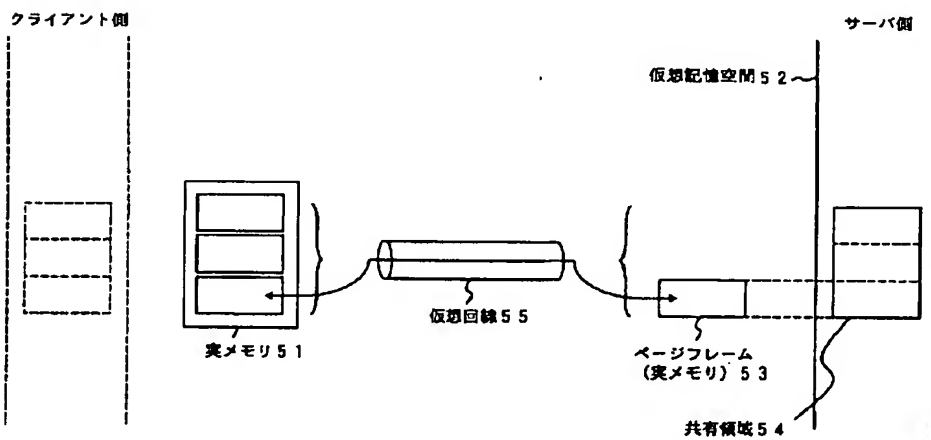
【図 7】



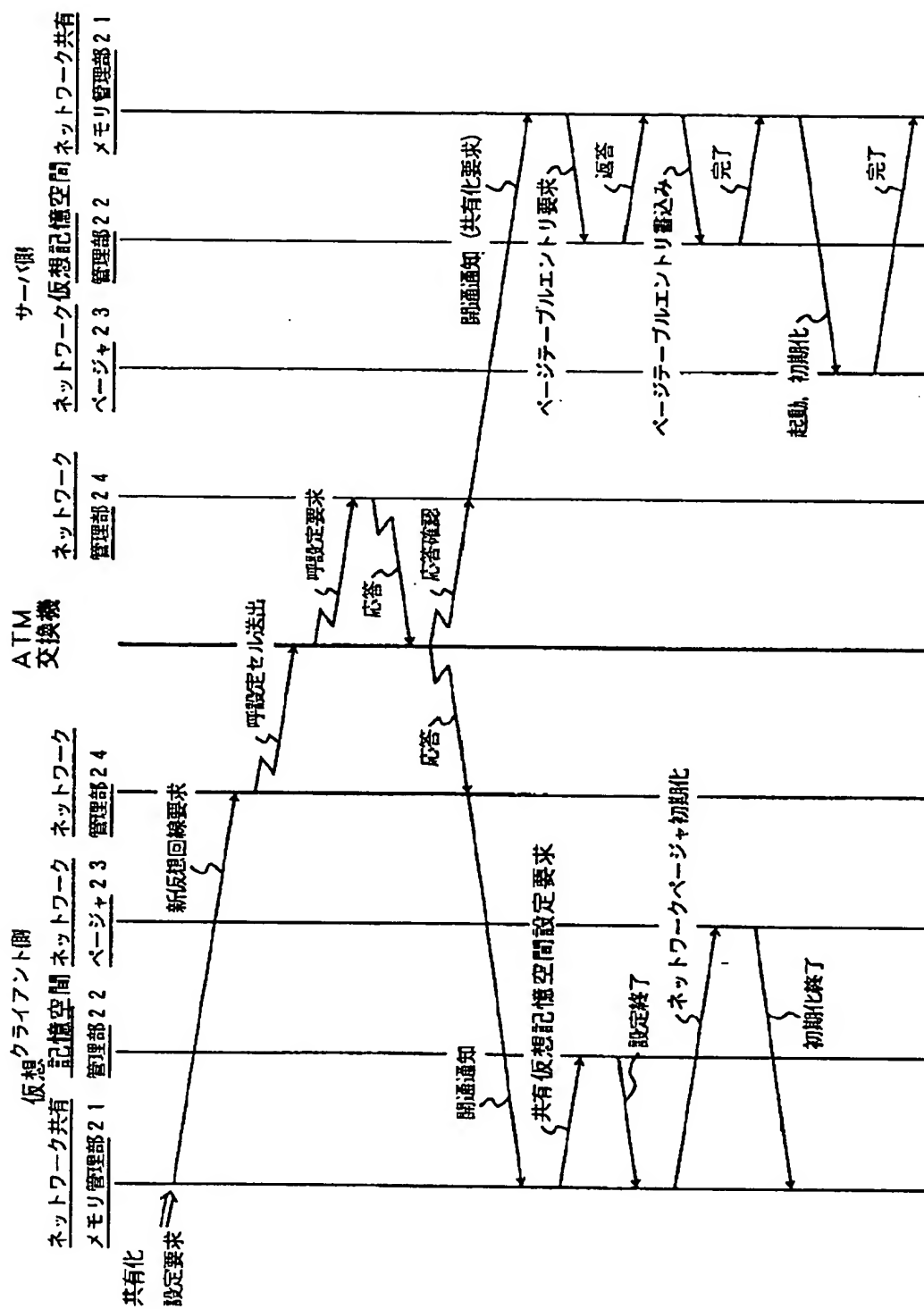
【図 1 4】



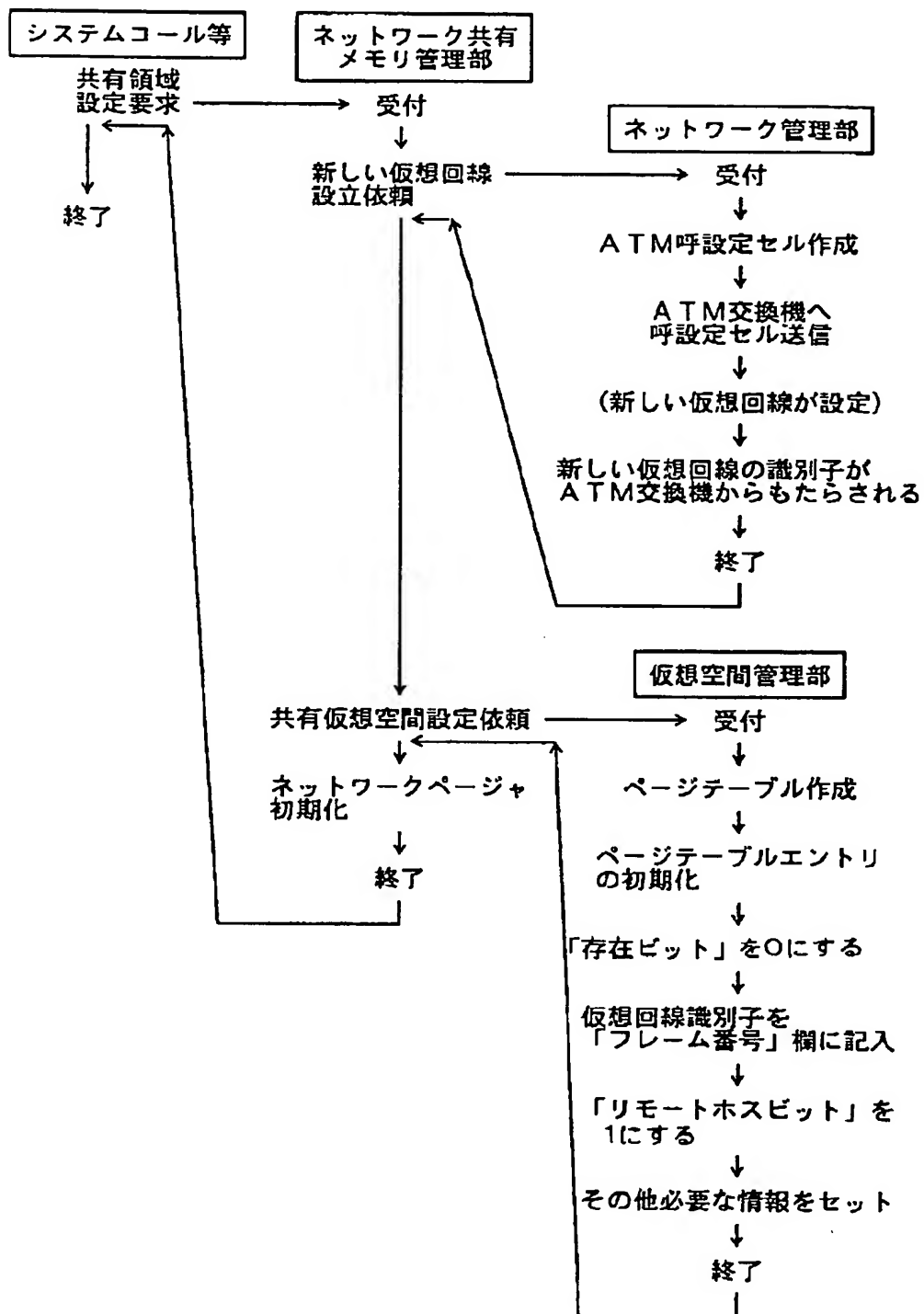
【図 1 5】



【図8】

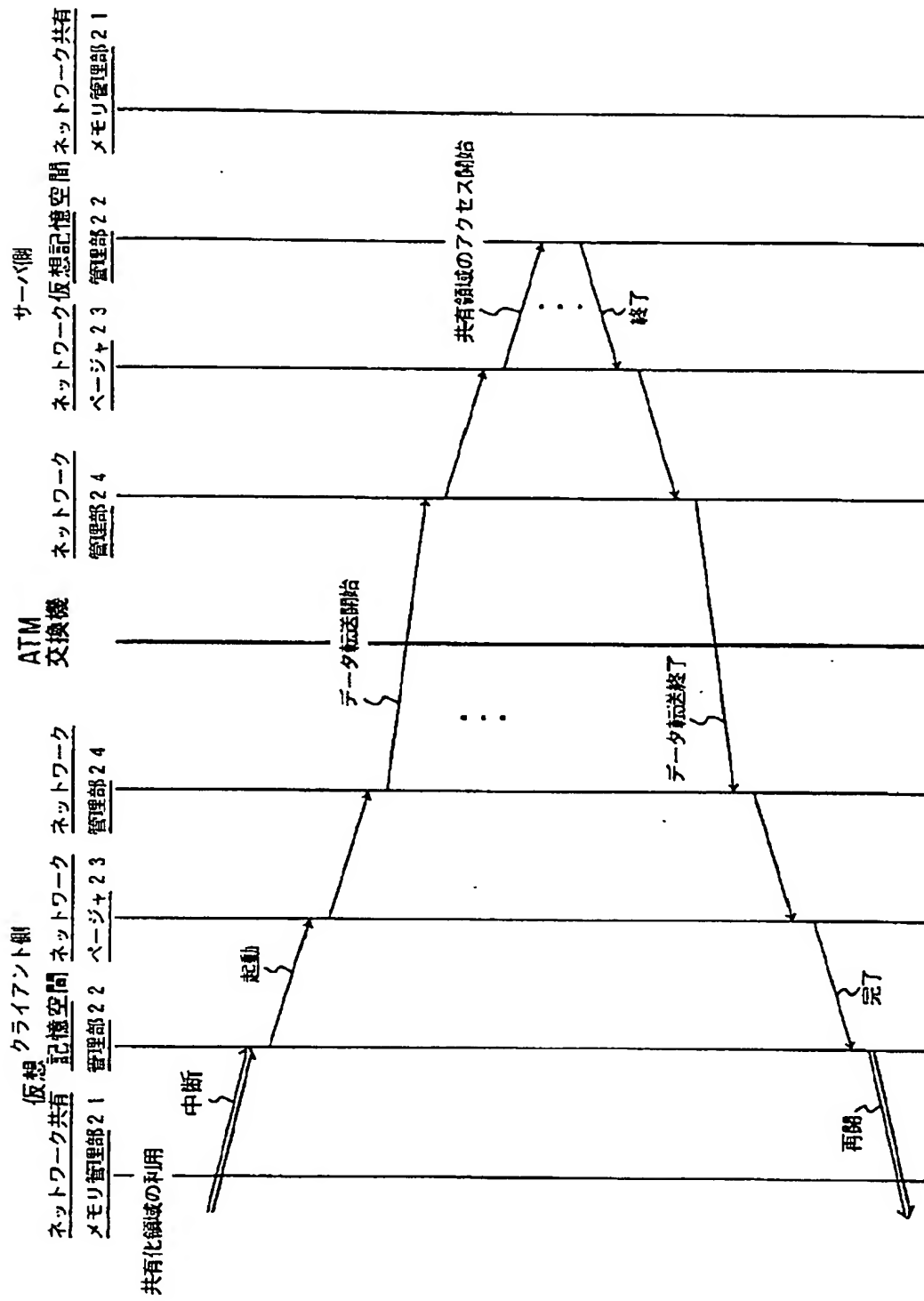


【図9】

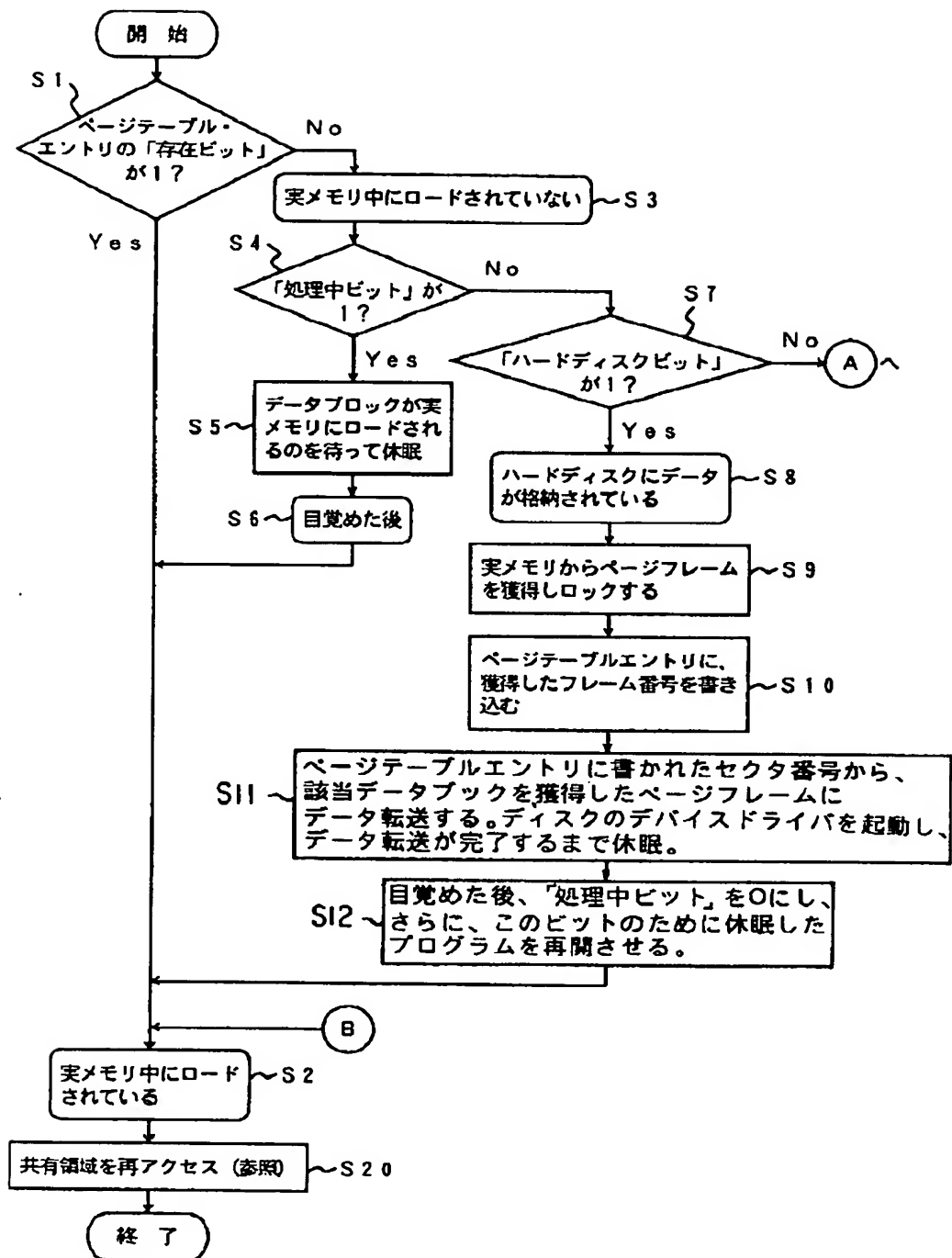


[illegible]

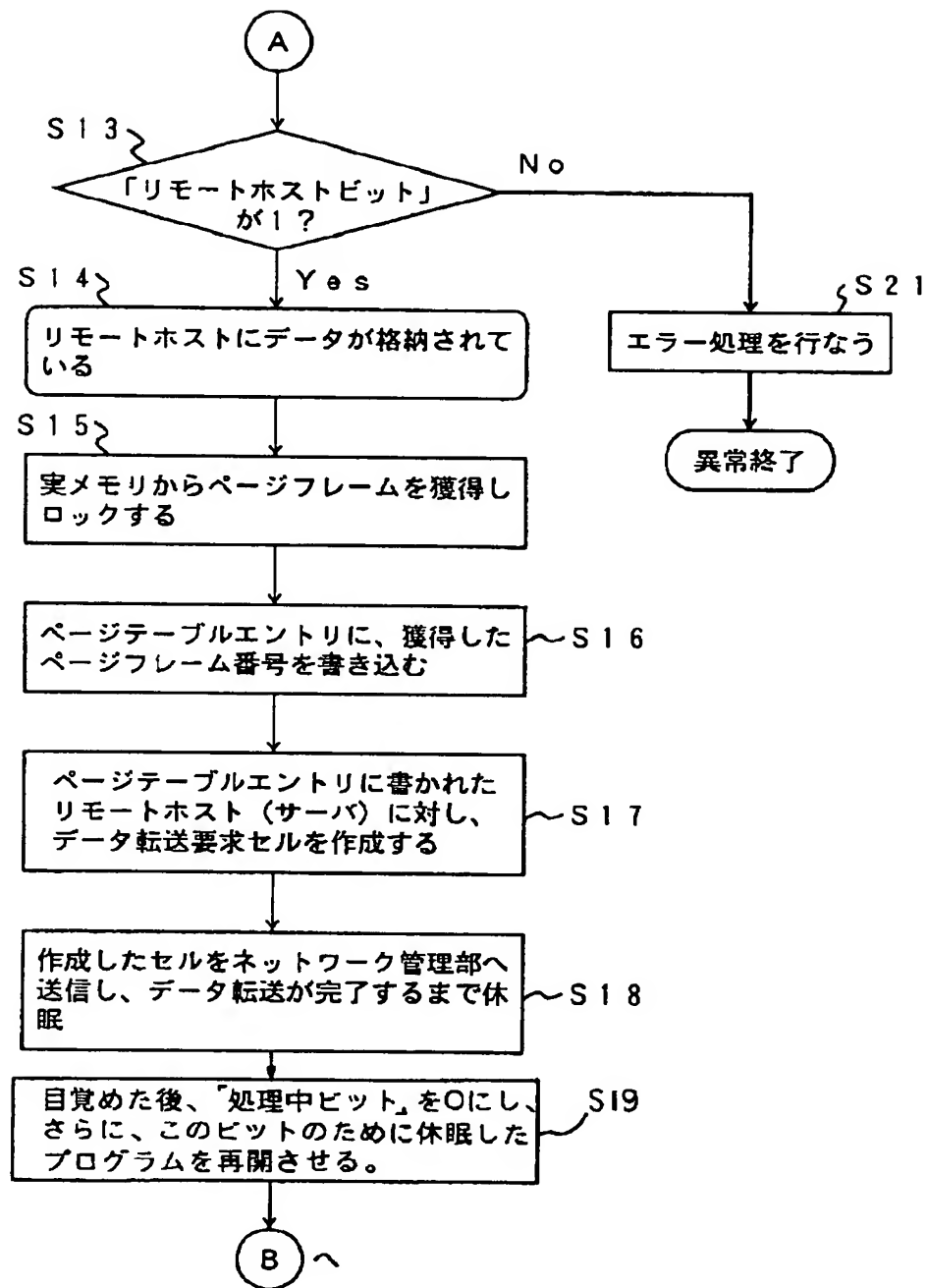
【図 1 1】



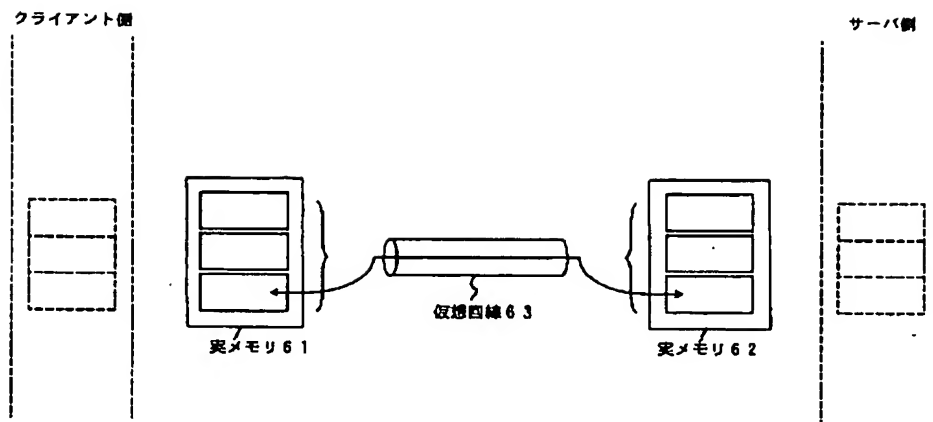
【図12】



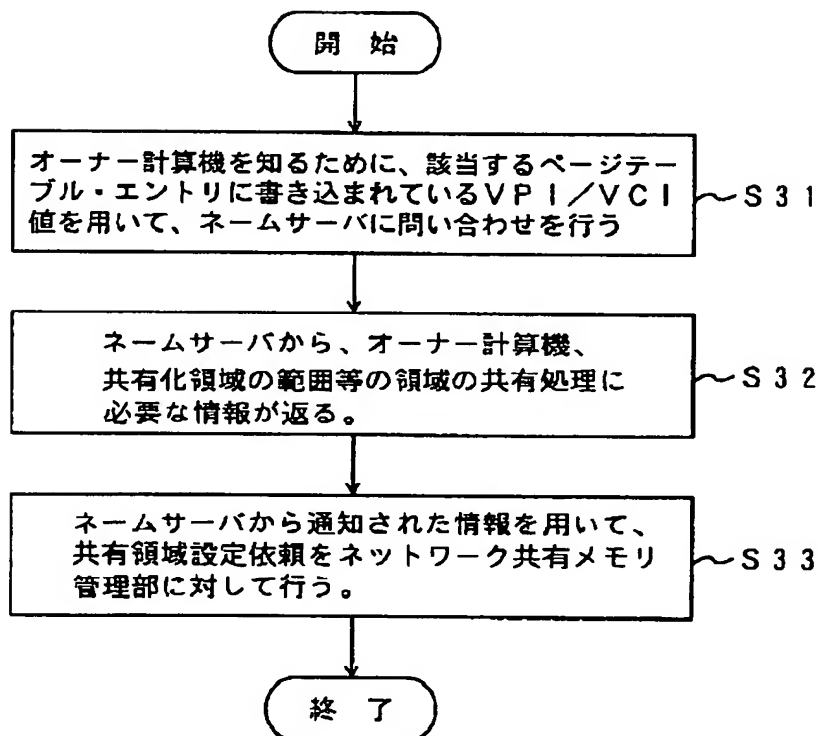
【図13】



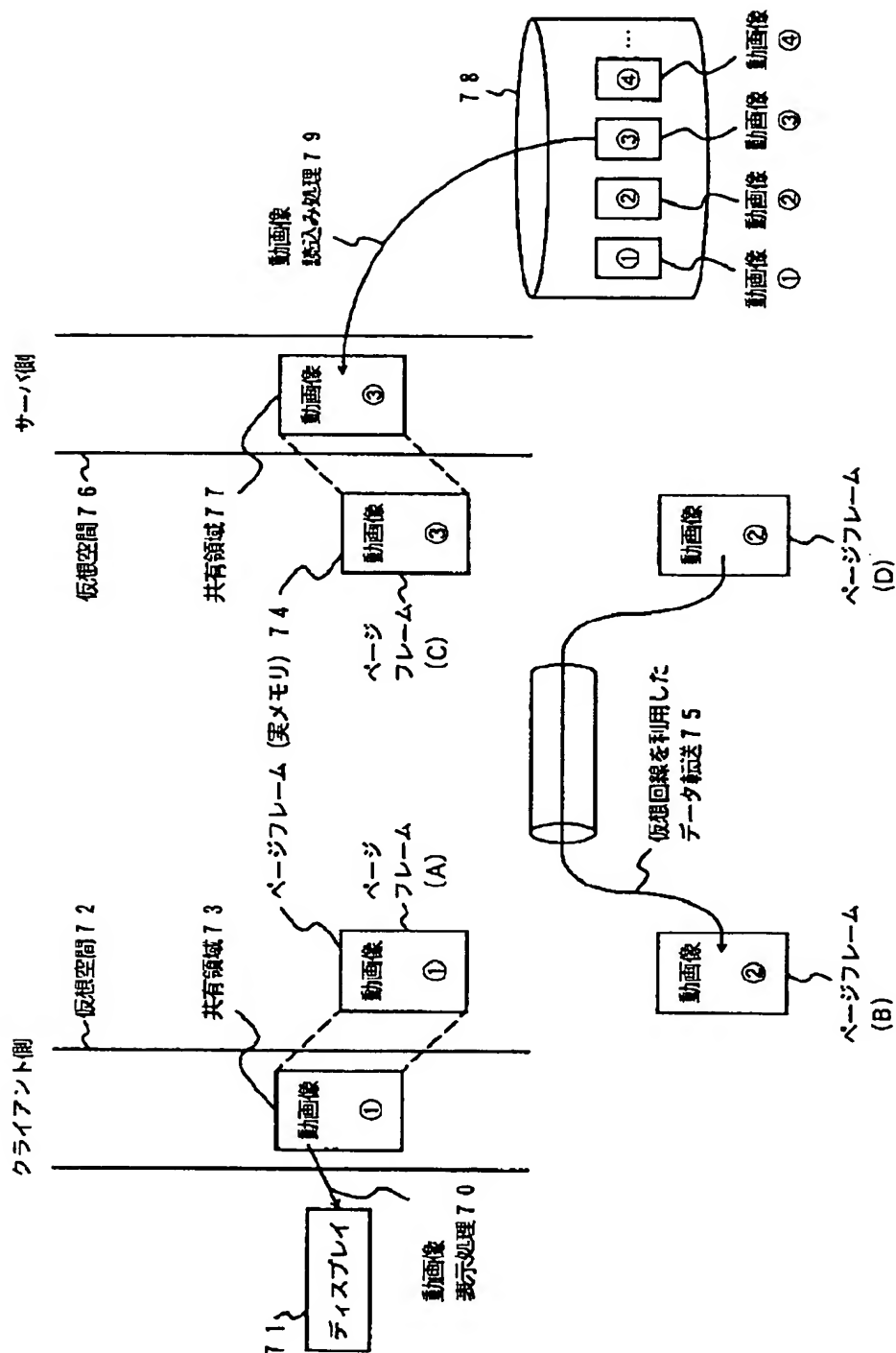
【図 1 6】



【図 1 9】



【図 17】



【図 18】

